KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Combining anatomical and spectral information to enhance MRSI resolution and quantification

Application to Multiple Sclerosis

Saurabh Jain

Supervisors:
Prof. dr. ir. F. Maes
Prof. dr. ir. S. Van Huffel
Dr. ir. D. Smeets

Dissertation presented in partial
fulfillment of the requirements for the
degree of Doctor of Engineering
Science (PhD): Electrical Engineering

May 2017

Combining anatomical and spectral information to enhance MRSI resolution and quantification

Application to Multiple Sclerosis

Saurabh JAIN

Examination committee:
Prof. dr. ir. C. Vandecasteele, chair
Prof. dr. ir. F. Maes, supervisor
Prof. dr. ir. S. Van Huffel, co-supervisor
Dr. ir. D. Smeets, co-supervisor
Prof. dr. P. Dupont
Prof. dr. U. Himmelreich
Prof. dr. B. Dubois
Prof. dr. ir. D. Vandermeulen
Dr. D. M. Sima
Prof. dr. S. Williams (Univ. of Manchester)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

May 2017

Uitgegeven in eigen beheer, Saurabh Jain, Kasteelpark Arenberg 10 box 2441, B-3001 Leuven (Belgium)

Preface

First of all, I would like to take the opportunity to sincerely thank Dirk Smeets. He is the first person I know in Belgium and my first contact at ico**metrix**. Starting from my first internship at ico**metrix** in 2012, I am his fan. He is a great supervisor, colleague and friend and has always inspired me to think one step ahead! From technical skills like theoretical formulation to soft skills like project management, I have learnt a lot from him. He is one of the first persons I can call, if I need help. I can write another thesis if I go in details, so I just want to thank you here Dirk for having faith in me and for your support, guidance and everything!

Diana Sima, I don't think I can ever thank you enough but still I will try! You are a 'superwoman', who can handle multitasking with efficiency! You are like a mother who is always ready to listen and help in all possible manners you can! You are like a buddy whose door I could knock anytime. You have helped me a lot in improving my technical writing skills, so if the reader enjoys reading this thesis, the credit goes to you!

I would like to thank my supervisor prof. Frederik Maes and my co-supervisor prof. Sabine Van Huffel for all the critical inputs throughout my PhD. Your inputs have always motivated me to improve my work. Thank you for having me as your student.

Thank you Annemie Ribbens for helping me in understanding the registration and segmentation concepts. You were always there to answer my stupid questions! Sorry that I nagged you sometimes by hiding your 'cola' from you ;-)

Thank you Dirk Loeckx and Wim Van Hecke for giving me the opportunity to work at ico**metrix**, the best place in the world I know to work in! Thanks Dirk Loeckx, I know it was not easy for you to hire me being a non-European, and for all the help so far during my stay in Belgium.

Now comes the amazing team of ico**metrix** that has converted this thesis into a real clinical product. I would like to thank Vasilis, Michiel, Anke, Lene, Ruben, Thijs, Jan, Felix, Arne, Thibo, Vân, Rien, Eloy, Maryna, Eline, Francisco, Karen, Hogne, Femke, Frauke, Jonas, Nataliya, Veronica and Len. Thank you all for the help and support!

Last but not the least, I would like to thank my family: Maa, Papa, Bhai, Bhabhi, my wife (Neha), Samiksha, Aagam, Ankit and Rashi! Without all of you, this journey would have not been possible!

Saurabh Jain
Leuven, May 2017

Abstract

Combining anatomical and spectral information to enhance MRSI resolution and quantification: Application to Multiple Sclerosis

Multiple sclerosis is a progressive autoimmune disease that affects young adults. Magnetic resonance (MR) imaging has become an integral part in monitoring multiple sclerosis disease. Conventional MR imaging sequences such as fluid attenuated inversion recovery imaging have high spatial resolution, and can visualise the presence of focal white matter brain lesions in multiple sclerosis disease. Manual delineation of these lesions on conventional MR images is time consuming and suffers from intra and inter-rater variability. Among the advanced MR imaging techniques, MR spectroscopic imaging can offer complementary information on lesion characterisation compared to conventional MR images. However, MR spectroscopic images have low spatial resolution. Therefore, the aim of this thesis is to automatically segment multiple sclerosis lesions on conventional MR images and use the information from high-resolution conventional MR images to enhance the resolution of MR spectroscopic images.

Automatic single time point lesion segmentation is performed using T1-weighted and fluid attenuated inversion recovery MR images where the brain is segmented into grey matter, white matter, cerebrospinal fluid and lesions using a probabilistic framework optimised by the expectation-maximisation technique. Then, a patch-based super-resolution method is used to increase the spatial resolution of metabolite maps computed from MR spectroscopic imaging. The patch-based super-resolution method uses high-resolution T1-weighted and fluid attenuated inversion recovery images together with the brain segmentations to regularise the super-resolution process. Finally, we extend the single time point lesion segmentation idea to two time points lesion segmentation. The two time points lesion segmentation is performed using T1-weighted and fluid attenuated inversion recovery MR images of the two time points optimised using a joint expectation-maximisation algorithm.

Validation of lesion segmentation methods on clinical datasets shows that they are accurate and consistent in segmenting multiple sclerosis lesions. Moreover, we study the association of a clinical biomarker – the Expanded Disability Status Scale score – and a MR imaging biomarker – new and enlarging lesion volume – and we observe that in majority of the patients, the new and enlarging lesion volume volume correlates better with Expanded Disability Status Scale score's evolution at patient level than at a group level. This study also shows the applicability of our method on multi-centre datasets without re-tuning or re-training, which makes it useful for the clinical use. Validation of the super-resolution method on synthetic and real images shows that our method preserves tissue contrast and structural information; and matches well with the trend of acquired high-resolution MR spectroscopic images. We analyse N-acetyl aspartate and myo-Inositol metabolites concentration in lesions and in the surrounding white matter. N-acetyl aspartate metabolite concentration in lesions is found to be lower compared to the surrounding white matter, and an opposite trend is observed for the myo-Inositol metabolite concentration.

From this research we conclude that the developed multiple sclerosis lesion segmentation methods, through their robustness and automation, could bring an added value to the clinical routine evaluation of multiple sclerosis patients. Also, the patch-based high-resolution MR spectroscopic images, through its tissue contrast conservation, can offer better lesion characterisation.

Samenvatting

Combinatie van anatomische en spectroscopische informatie om MRSI resolutie en kwantificering te verbeteren: Toepassing op Multiple Sclerose

Multiple sclerose is een progressieve auto-immuunziekte die jonge volwassenen treft. Magnetische resonantie (MR) is een integraal onderdeel bij het toezicht op multiple sclerose ziekte. Conventionele MRI sequenties hebben een hoge resolutie en kunnen de aanwezigheid van focale witte stof hersenletsels bij multiple sclerose ziekte visualiseren. Manuele aflijningen van deze laesies op conventionele MR beelden is echter tijdrovend en lijdt aan intra- en inter-rater variabiliteit. Onder de geavanceerde MRI-technieken, kan MR spectroscopische beeldvorming aanvullende informatie geven over laesie karakterisering in vergelijking met conventionele MR-beelden. Maar MR spectroscopische beelden hebben een lage resolutie. Daarom is het doel van dit proefschrift om de segmentatie van multiple sclerose laesies op conventionele MR-beelden te automatiseren en de informatie van hoge-resolutie conventionele MR-beelden te gebruiken om de resolutie van MR spectroscopische beelden te verbeteren.

Automatische enkel tijdstip laesie segmentatie wordt uitgevoerd met behulp van T1-gewogen en FLAIR MR-beelden waarbij de hersenen worden gesegmenteerd in grijze stof, witte stof, cerebrospinale vloeistof en laesies met behulp van een probabilistisch model geoptimaliseerd door de verwachting-maximalisatie techniek. Vervolgens wordt een patch-gebaseerde superresolutie methode ontwikkeld die de resolutie van metaboliet beelden, berekend uit MR spectroscopische beeldvorming, verhogen. De patch-gebaseerde super-resolutie methode maakt gebruik van hoge resolutie T1-gewogen en FLAIR beelden samen met de hersensegmentatie om het super-resolutie proces te regulariseren. Tot slot, wordt de enkel tijdstip laesie segmentatie idee uitgebreid om twee tijdstippen laesie segmentatie te bekomen. De twee tijdstippen laesie segmentatie wordt uitgevoerd op basis van de T1-gewogen en FLAIR MR beelden van de twee tijdstippen, en het model wordt geoptimaliseerd met behulp van een gekoppeld verwachting-maximalisatie algoritme.

Validatie van laesie segmentatie methoden op klinische datasets laat zien dat ze accuraat en consistent zijn in het segmenteren van multiple sclerose laesies. Bovendien onderzoeken we de associatie van een klinische biomerker - de Expanded Disability Status Scale score - en een MRI biomerker - volume van nieuwe en groeiende laesies - en we zien dat in de meerderheid van de gevallen het volume van nieuw en groeiende letsels beter correleert met de evolutie van de Expanded Disability Status Scale score op patientniveau dan op groepsniveau. De studie heeft ook aangetoond dat onze methode toepasbaar is op datasets van verschillende centra, zonder calibratie. Dit maakt de methode aantrekkelijk voor klinisch gebruik. Validatie van de super-resolutie methode van synthetische en echte beelden toont aan dat onze methode weefselcontrast en structurele informatie behoudt, en past goed bij de trend van hoge-resolutie MR spectroscopische beelden. We analyseren N-acetyl aspartaat en myo-Inositol metabolieten concentratie in laesies en in de omliggende witte stof. N-acetyl aspartaat metaboliet concentratie laesies blijken lager te zijn dan de omringende witte stof, en een tegengestelde trend wordt waargenomen voor de myo-Inositol metaboliet concentratie.

Uit dit onderzoek kunnen we concluderen dat de ontwikkelde multiple sclerose laesie segmentatie methodes, door hun robuustheid en automatisering, een toegevoegde waarde kan betekenen voor de klinische routine evaluatie van multiple sclerose-patiënten. De patch-gebaseerde high-resolution MR spectroscopische beelden, door middel van het behoud weefselcontrast, kunnen ook bijdragen tot een beter karakterisering van laesies.

Abbreviations

BBB	Blood brain barrier
BET	Brain extraction tool
BS	B-splines
Cho	Choline
CNS	Central nervous system
Cr	Creatine
CSF	Cerebrospinal fluid
DIS	Dissemination in space
DIT	Dissemination in time
DTI	Diffusion tensor imaging
EDSS	Expanded disability status scale
EM	Expectation-maximisation
FLAIR	Fluid attenuated inversion recovery
Glx	Glutamate and glutamine
GM	Grey matter
GMM	Gaussian mixture model
ICAM-1	Intracellular adhesion molecule 1
ICC	Inter-class correlation
LFPR	Lesion-wise false positive rate
LIN	Linear interpolation
LRSR	Linear regression based super-resolution
LST	Lesion segmentation tool
LTPR	Lesion-wise true positive rate

MAP	Maximum a posteriori
MRF	Markov random field
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
MRSI	Magnetic resonance spectroscopic imaging
MS	Multiple sclerosis
MTR	Magnetisation transfer ratio
myo-Ins	myo-Inositol
NAA	N-acetyl aspartate
NEL	New and enlarging lesions
NN	Nearest neighbour
PBSR	Patch-based super-resolution
PPMS	Primary progressive multiple sclerosis
PRESS	Point-resolved spectroscopy
PRMS	Progressive relapsing multiple sclerosis
PSF	Point spread function
RRMS	Relapsing-remitting multiple sclerosis
SNR	Signal to noise ratio
SPMS	Secondary progressive multiple sclerosis
SR	Super-resolution
SSIM	Structural similarity index
STEAM	Stimulated echo acquisition mode
SVS	Single voxel spectroscopy
VCAM-1	Vascular cell adhesion molecule 1
WM	White matter

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Multiple Sclerosis

Multiple Sclerosis (MS) is an inflammatory autoimmune disorder that results in a neuro-degeneration of the central nervous system (CNS). It is a progressive disease that disseminates in both space and time in the brain and spinal cord of adults. MS symptoms can vary from patient to patient; however, the most common symptoms include (1) motor deficits such as: ataxia (lack of coordination of muscle movements), weakness, tremor (somewhat rhythmic relaxation and contraction), and spasticity (muscle stiffness), (2) fatigue: inability to generate force capacity to carry out normal day to day work, (3) vision problems: temporary vision loss, eye pain, colour blindness, (4) abnormal sensations: numbness or tingling in different parts of the body such as arms and legs, (5) sexual problems, (6) bladder problems [75, 184].

Cause and pathological physiology
The exact cause of MS is unknown. However, it is associated with the destruction of the myelin sheets in the CNS. In order to describe what causes this destruction, it is important to introduce the biological terminology. Glial cells are the non-neural cells found in the CNS that mainly consists of: oligodendrocytes, microglia and astrocyte cells. An oligodendrocyte is a myelinating cell in the CNS that synthesises the myelin sheet around a neuron that acts as an insulator and helps a neuron in communicating with the other neurons. A microglial cell

---

Sections 1.4, 1.5 and 1.6 have been adapted from: Sima D.M., Loeckx D., Smeets D., Jain S., Parizel P.M., and Van Hecke W. Use Case I: Imaging Biomarkers in Neurological Disease. Focus on Multiple Sclerosis, In *Imaging Biomarkers: Development and Clinical Integration*, Eds. Martí-Bonmatí, L. and Alberich-Bayarri, A., Springer International Publishing, 2016, 169-180.

is a cleaning agent that constantly scavenges for dead/damaged neurons and infectious agents in the CNS. An astrocyte cell provides structural support to the neurons and helps in neuronal growth and repair. Moreover, it also provides biochemical support to endothelial cells that consist of vascular cell adhesion molecule 1 (VCAM-1) and intracellular adhesion molecule 1 (ICAM-1). These molecules control the permeability of the blood brain barrier (BBB), which allows only certain types of substances (e.g. monocytes: a type of white blood cells) to enter the CNS. Figure 1.1-A provides a visual presentation of this scenario. In case of MS, due to some unknown factor, a monocyte is wrongly programmed to induce immune reaction to the MS specific antigen - a substance that causes the immune system to produce antibodies against it. This monocyte enters the CNS system through BBB, activating the glial cells to react against the MS specific antigen. Microglial cells together with other immune cells then attack the oligodendrocyte cells together with the myelin sheets, resulting in demyelination. Moreover, attack on the astrocyte cell results in an increase in the permeability of the BBB resulting in more and more monocytes entering the CNS (see Figure 1.1-B)[128].

Worldwide prevalence and types
According to the 2013 survey of the global MS atlas, 2.3 million people are affected worldwide. Although MS is present throughout the world, it is more commonly found in the European countries, Canada and United States of America (see figure 1.2). In general, MS is twice as common in women than men, and the patient life expectancy after MS onset is at least 25 years [109]. The four main clinical forms of MS are relapsing-remitting MS (RRMS), secondary progressive MS (SPMS), primary progressive MS (PPMS) and progressive relapsing MS (PRMS). The most general form of MS is RRMS ($\sim$85%) in which the patient experiences relapses (when symptoms flare up) that may or may not recover completely. Approximately 80% of the RRMS patients enter the SPMS phase. Unlike RRMS, SPMS patients experience permanent neurological disability along with occasional relapses. The remaining 15% of the patients who do not have RRMS form at the disease onset, have progressive form of MS: PPMS or PRMS. In PPMS, the patient has progressive early phase neurological disability and no early phase relapses. In very few rare cases, some patients have early phase progressive neurological disability with early phase relapses, which is known as the PRMS form of MS [113].

1.2 Clinical diagnosis and treatment of MS

The initial symptoms of MS disease often overlap with other diseases such as acute and recurrent disseminated encephalomyelitis, and thus makes its

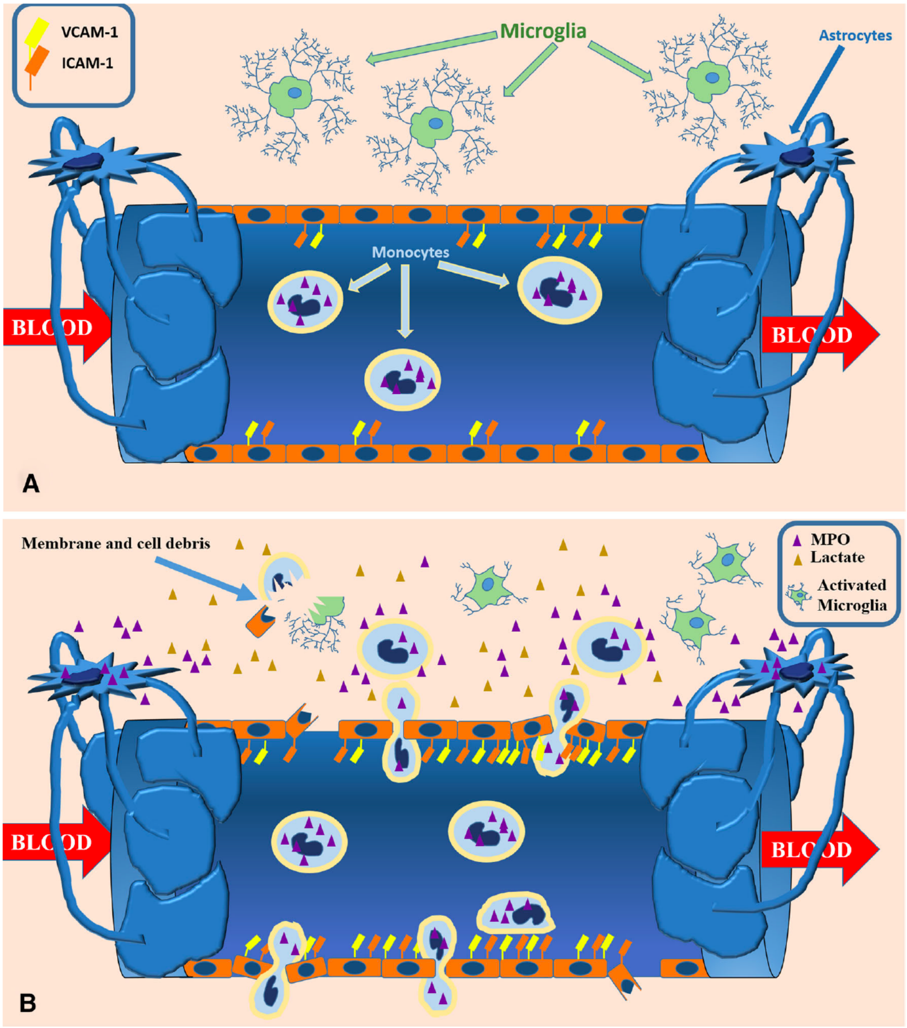


Figure 1.1: Normal working of the immune system (A) followed by its alteration in MS (B). (Figure 1 from [128]; original caption: *Main neuroinflammatory mechanisms probed by MRI. Modifications of the neurovascular unit include alterations of the BBB permeability, associated with overexpression of adhesion molecules,which induce blood-borne monocytes to arrest and crawl along the endothelium, crossing it by diapedesis along loosened intercellular junctions. Machrophage and microglia activation results in increased MPO activity and lactate accumulation in the interstitium, along with the presence of cellular debris resulting from cell damage due to oxidative and excitotoxic phenomena.* Article [128] is distributed under the terms of the Creative Commons Attribution 4.0 International License (`http://creativecommons.org/licenses/by/4.0`), by Springer Publications, Ltd.: Clinical and translational imaging, 6(5):475-489, copyright © 2015.)

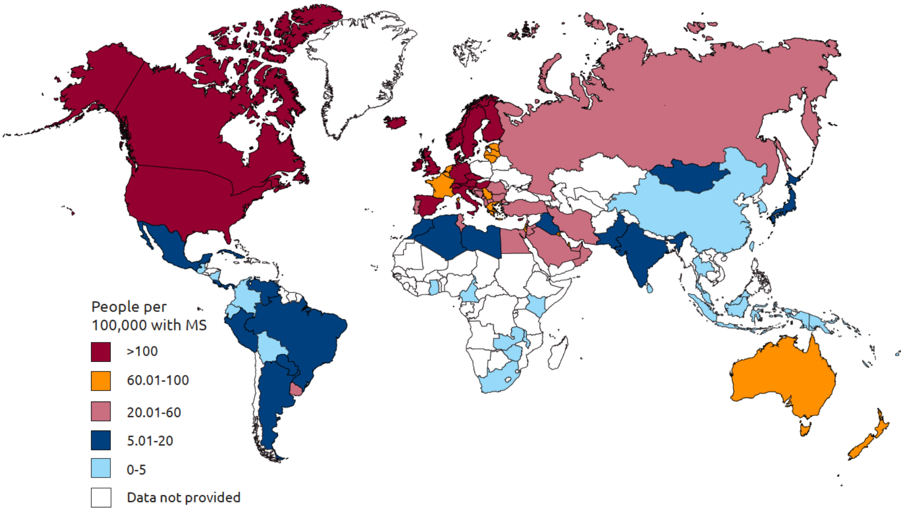


Figure 1.2: Worldwide prevalence of MS. (Figure from [109]; original caption: *PREVALENCE BY COUNTRY (2013)*. Reprinted with permission from Multiple Sclerosis International Federation Atlas of MS 2013, www.atlasofms.org, copyright © 2013.)

diagnosis very challenging. The concept of disease dissemination in space and time is very important in MS diagnosis as it separates MS from other diseases. According to the revised McDonald criteria for diagnosing MS, if the patient suffers two clear attacks (relapses; exacerbation) at least 1 month apart (dissemination in time) with objective clinical evidence (classification of presenting symptoms and signs as either monofocal or multifocal) of at least two lesions disseminated in space are sufficient criteria for MS diagnosis solely based on clinical grounds [119]. Clinical evolution of MS is characterised by both motor and cognitive degradation [24, 129]. Pathological changes in the normal appearing white and grey matter are better correlated with progressive cognitive deficits than with visual, sensory, and motor symptoms [12, 66, 145]. Based on the clinical symptoms, the most common clinical biomarkers that are used to monitor disease progression include annualised relapse rate and Expanded Disability Status Scale (EDSS). EDSS score measures the patient disability on a scale from 0 to 10, with higher scores signifying worse patient condition [97].

In current clinical practice of MS, the assessments of the extent of brain damage and drug efficacy are based on visual inspection of MR images by expert neurologists and neuroradiologists, who evaluate the presence and distribution of focal white matter lesions. A huge body of research has been devoted to white matter lesions, since they are considered to be a distinctive feature for the disease diagnosis, progression and treatment.

There is no cure for MS as the underlying cause is still unknown. However, treatments that can favourably affect MS are generally based on immune modulation or symptomatic improvements. Immune modulation therapy decreases the activation and proliferation of monocytes into the CNS. Examples include interferon-beta, fingolimod, glatiramer acetate, teriflunomide, nataluzimab, alemtuzumab etc. Symptomatic therapy on the other hand does not change the underlying disease but pacify the different symptoms' intensity. Examples include amantadine for fatigue etc [75].

In the past, reducing the number of relapses and its consequences was the sole aim of the immune modulation therapies. Recent studies have shown that both relapses and ongoing MRI activity are associated with the disease activity [44, 13]. As a result, there is a shift in the primary aim of the treatment being 'no evidence of disease activity' (NEDA)[69]. NEDA refers to 'freedom from the disease activity'. It is a new and evolving concept that aims to monitor treatment of RRMS patients to reach a state where they have (a) no relapses (b) no new and active lesions on their MRI scans (c) no increase in their disability. The revised NEDA criteria [88] also include brain volume change (also known as brain atrophy) to improve the predictive value of the treatment. The inclusion of MRI based measures in the NEDA criteria shows the growing importance of MRI in monitoring the disease activity.

1.3 MRI in the diagnosis and monitoring of MS

White matter lesions (also known as 'plaques') are considered to be a hallmark of the disease and can be visualised with several MRI sequences [149]:

- T1-weighted MR images: chronic stage lesions with axonal destruction and irreversible damage appear as dark spots ('black holes'), compared to the surrounding white matter (WM) tissue intensities (see Figure 1.3).
- Gadolinium-enhanced T1-weighted MR images: 'active' lesions taking up contrast material, and indicating inflammation and breakdown of the blood-brain-barrier.
- T2-weighted MR images, fluid attenuated inversion recovery (FLAIR) MR images and proton density (PD) MR images: on these imaging sequences with a long repetition time (TR), lesions appear as hyper-intense spots compared to the surrounding brain parenchyma (see Figure 1.3).

MS diagnosis using MRI

Diagnosing a patient solely on clinical grounds is often very challenging specially due to the difficulty in classifying the patient symptoms as disease dissemination in space. Therefore, the revised McDonald criteria also includes the guidelines on using the MRI if the patient cannot be solely diagnosed on clinical grounds. The revised McDonald MRI criterion for dissemination in space criteria includes: (a) at least one T2 lesion in at least 2 out of 4 CNS areas: spinal cord, infratentorial, periventricular or juxtacortical. Dissemination in time includes: (a) a new T2 and/or gadolinium-enhancing lesion on follow-up scan, irrespective of its timing with reference to the baseline scan; (b) simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time [119].

MS monitoring using MRI

In a clinical scenario, repeated MRI scans are visually inspected to detect and monitor disease activity [181]. The most commonly used MRI sequences include Gadolinium-enhanced T1-weighted MR images and T2-weighted/FLAIR MR images. Presence of Gadolinium-enhanced lesions indicate ongoing disease activity, since only new lesions (under 6 weeks old) enhance (see Figure 1.4). New lesions on a follow-up T2-weighted/FLAIR image suggest an inflammatory reaction that has developed between two MRI scans and provide information on the disease activity and disease progression [60, 159].

As the correlation between established MRI biomarkers for monitoring the MS disease activity and clinical outcome measures is low, there is a need to further investigate other possible MRI biomarkers [117]. Towards this end, several studies have shown that a high lesion load and the location of MS lesions have a predictive value for the development of clinical disability [57, 168]. ‘Lesion load’, is defined as the total volume of lesions in the brain, is one of the most important biomarkers in MS for monitoring the disease activity. Often, a distinction is made between T2-lesions (*i.e.,* lesions that appear hyperintense on T2-weighted or FLAIR images), T1-lesions (*i.e.,* lesions that appear hypointense on T1-weighted images, the so-called black holes), as well as contrast-enhancing lesions. In addition to the lesion load, brain volumetry [67], more specifically, total brain atrophy [13] and, in particular, grey matter (GM) atrophy [65], are currently considered to be essential biomarkers for monitoring the MS disease activity, since they are correlated with the speed of disease progression (see Figure 1.5, table 1.1). Thus, apart from the detection of lesions, quantification of brain volumes and atrophy rates are increasingly important in the management of patients with MS. Combined clinical and MRI course of MS is shown in figure 1.6. Moreover, the inclusion of brain atrophy, and new and active lesions in the NEDA criteria shows the growing importance of MRI in monitoring the disease activity.

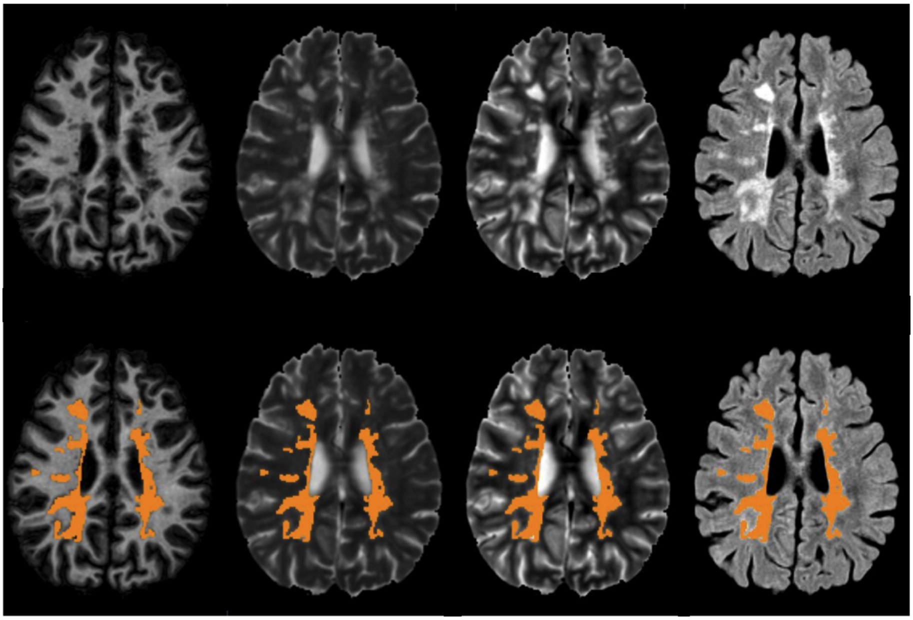


Figure 1.3: Top, from left to right: T1-weighted, T2-weighted, proton density and FLAIR images, obtained in an MS patient. Bottom: same images overlaid with expert manual delineation of MS lesions. (Data from the 'MS longitudinal lesion segmentation challenge', ISBI 2015; training subject 02, time point 01.)

MRI in monitoring and predicting treatment response in MS

Several clinical trials include standard MRI biomarkers (lesion load, active T2 lesions, contrast-enhancing lesions) as secondary outcome measures while investigating the clinical efficacy of a treatment in MS. A meta-analysis of several treatment trials have shown that the effect of treatment on MRI lesions in a phase 2 trial reliably predicts the effect on clinical relapse activity in a consecutive phase 3 trial with the same substance [158]. Translation of clinical trials outcome is often very difficult to apply in clinical routine. Clinical studies have demonstrated that the appearance of a new T2 lesion or contrast-enhancing lesion during the first year of a treatment (mostly on interferon-β) correlates with disability progression [181, 125, 132].

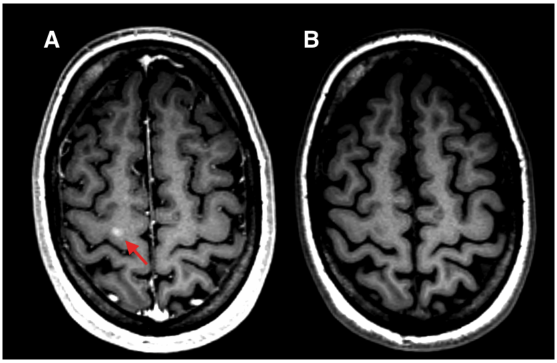


Figure 1.4: Active lesion (A) appearance in the post-contrast T1-weighted MRI (red arrow), (B) absence in the pre-contrast T1-weighted MRI obtained in a patient with relapsing-remitting multiple sclerosis.

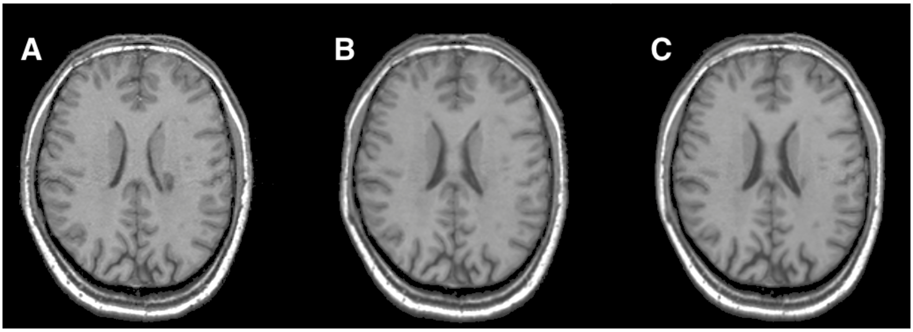


Figure 1.5: Brain atrophy in an MS patient: (A) Baseline scan. (B and C) Regular scans taken at an interval of 4 years. Progressive neurodegeneration can be seen by the enlargement of the ventricles, which signify brain volume loss over time.

1.4 MRI acquisition requirements

Widespread application of MR imaging biomarkers is hampered by issues such as: non-standardised imaging protocols, imaging artefacts, lack of normative data and manual segmentations to quantify values in clinical practice. In order to mitigate such issues and promote MR imaging in MS clinical practice, the MAGNIMS Study Group published guidelines for the use of MRI in MS diagnosis [137], as well as recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal MS studies [179, 163]. The group recommends that ‘images should be acquired using 3D pulse sequences, with near-isotropic spatial resolution and multiple image contrasts to allow more

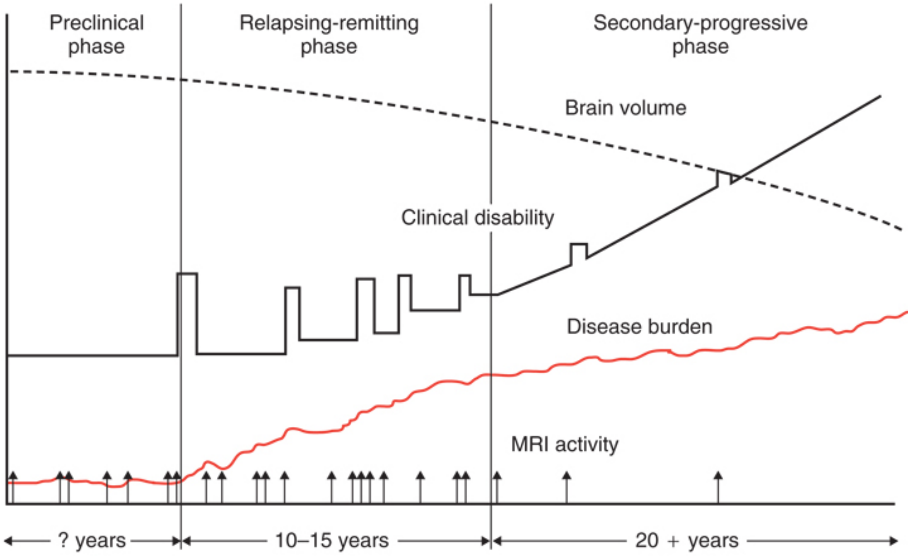


Figure 1.6: Combined clinical and MRI course of MS. Figure from [58]; original caption: *Typical clinical and MRI course of multiple sclerosis. MRI activity (vertical arrows) indicates an inflammatory process as measured on brain MRI by gadolinium enhancement or new T2 hyperintense brain lesion. MRI activity is typically more frequent than clinical relapses (spikes in clinical disability), which indicates that more disease activity is taking place than is clinically apparent. Loss of brain volume and increase in disease burden (total volume of lesions), both measured on MRI, indicate permanent tissue damage, which is present early in the disease and gradually progresses over time.* Reprinted with permission from Fox RJ, Cohen JA. Multiple sclerosis: the importance of early recognition and treatment. Cleve Clin J Med 2001; 68:157-171. Copyright © 2001 Cleveland Clinic Foundation. All rights reserved.

Table 1.1: Evidence for the relevance of lesion and atrophy biomarkers in MS.

Biomarker	**Findings**	**Study population**	**Reference**
T2-lesions	annualised change: +0.25 ± 0.5 ml	10-year follow-up, RRMS	[68]
T1-lesions	annualised change: +0.20 ± 0.31 ml	Idem	Idem
T2-lesions	annualised change: +0.80 ml/year	20-year follow-up, RRMS	[57]
T2-lesions	correlated with disability progression	10-year follow-up	[120]
Total brain atrophy	correlated with disability progression	Idem	Idem
Total brain atrophy	CIS = -0.40% ± 0.47%, RRMS = -0.49% ± 0.65%, SPMS = -0.64% ± 0.68%, PPMS = -0.56% ± 0.55%	>1-year follow-up 963 subjects	[40]
GM atrophy	significant differences between progressive vs. non-progressive	10-year follow-up	[81]

GM = Grey matter, CIS = Clinically isolated syndrome, RRMS = Relapsing-remitting multiple sclerosis, SPMS = Secondary progressive multiple sclerosis, PPMS = Primary progressive multiple sclerosis.

comprehensive analyses of lesion load and atrophy, across time points. Image artefacts need special attention given their effects on image analysis results' [179]. Image artefacts interfering with MRI readings include radio-frequency (RF) intensity non-uniformity, phase-encode ghosting, signal wrap, and geometric distortion due to gradient non-uniformity, and B_0 inhomogeneity.

Investigators of the Canadian MRI Consensus Group further specify that 'a standardised MRI protocol is important during patient follow-up' [5]. The recommended brain MRI sequences are: 3D FLAIR (or axial + sagittal FLAIR), post-gadolinium T1, axial T2 and/or PD, obtained with a minimum MRI field strength of 1.5T, and a slice thickness: 1 mm for T1 and $\leq$ 3 mm for FLAIR with no gap; total head coverage should include the entire brain and brainstem.

MSCare, the (US) Consortium of Multiple Sclerosis Centers [169], proposes a standardised MRI protocol that they regularly update. The recommendations include (a) sufficient signal-to-noise ratio and resolution (in-plane resolution of 1 mm^2 with $\leq$ 3 mm slice thickness, no gap), (b) whole brain coverage, (c) usage of 3D T2-weighted FLAIR (sagittally acquired) or 3D T2-weighted or 2D diffusion weighted imaging sequences (for complete guidelines, see table 2 in [169]). These requirements are sufficient not only for visual assessment, but also for automated image analysis software, since most packages performing MRI-based brain segmentation, atrophy computations or lesion segmentation work either on single MR images or on a subset of multi-parametric images, simultaneously taken into account.

1.5 MRI analysis methods

1.5.1 Cross-sectional biomarkers

Brain volume computations
The volume of the whole brain, or volumes of brain structures, can be easily computed through brain segmentation techniques. Brain segmentation implies that the whole brain, its constituent tissue types, or individual brain substructures, can be identified based on MRI(s). A typical first step is 'brain extraction' or 'skull stripping', a preprocessing step that ensures that only brain tissue is transmitted to the segmentation pathway. Various brain extraction methods such as the brain extraction tool (BET) [152], brain surface extractor (BSE) [144], ROBEX [78], etc., are available. Approaches are diverse, including morphological, geometrical, image processing and modelling operations (hole filling, surface modelling, edge detection, intensity thresholding, atlas matching, deformable models, patch-based labelling, etc.). Moreover, the results of

individual brain extraction methods can be enhanced by applying hybrid techniques, thus combining results from several individual methods.

After the first step of brain extraction, the process of brain segmentation can be started. This is typically based on a probabilistic modelling of voxel intensities, exploiting the fact that different tissue types have different MR image characteristics. Recent literature provides an excellent overview of brain segmentation methods [42]. Well-known and validated examples include FAST [185], SIENAX [153], FreeSurfer [54]. Gaussian mixture models are popular; image intensities for each tissue type are modelled as a (sum of) Gaussian components. This modelling is usually performed using expectation-maximisation (EM), a well-known iterative parameter estimation algorithm. Spatial priors, serving as starting values and also as spatial constraints, can be obtained from appropriate brain atlases available in the literature [131]. The EM framework can be extended to intrinsically model some of the common distortions present in MR images, such as spatial inhomogeneity of image intensities known as bias field. Otherwise, such correction should be performed in pre-processing, *e.g.,* using methods such as N3 or N4ITK [150, 171]. The EM results are probabilistic, i.e., each voxel is assigned probability of belonging to each of the classes of interest (WM, GM, cerebrospinal fluid (CSF), etc.). These maps can be thresholded to obtain hard segmentations. Volumes in millilitres for each class can be computed either based on the hard or the fuzzy segmentation, by simply multiplying the sum of the tissue segmentation over all voxels by the voxel volume.

Lesion detection and volume estimation

Some automatic lesion segmentation methods belong to the family of supervised classification methods, for which a representative training dataset, including expert segmentation, is required in order to build a model that can be used on new patients for lesion segmentation. Depending on the features extracted from images (local gradient intensity, mean intensity, spatial information, etc.) and on the type of classifier (k-nearest neighbours, artificial neural networks, Bayesian learning, support vector machines, etc.), many variants have been proposed [63, 96, 91, 160]; see also [61, 108] for overviews of algorithms and software solutions). Although excellent results can be obtained with supervised classification on the training dataset, these methods have two disadvantages. The first difficulty lies in building a training dataset that encompasses MS lesions of all possible shapes, intensities and are heterogeneously distributed in the WM. The second non-trivial problem lies in pre-processing a new image (acquired on a different scanner than the one used for the training dataset), such that it matches the characteristics of the training dataset, *e.g.,* by intensity normalisation. In other words, supervised methods perform well only when the new image to be segmented is well represented in the training dataset.

Another family of methods is based on unsupervised classification and does not require training images. These methods are usually based on stochastic modelling of voxel intensity distribution. They perform brain segmentation into GM, WM and CSF (with or without lesion detection) and often rely on post-processing approaches in order to segment lesions (*e.g.,* lesion growing or pruning). The assumptions that are made in order to segment lesions have a great impact on the results. For instance, Van Leemput et al. [176] use three MR sequences (T1-, T2- and PD-weighted) to segment MS lesion as an outlier to the normal brain. In addition, it corrects for MR field inhomogeneities and incorporates spatial consistency using Markov Random Field. LST [142] detects FLAIR-hyper-intense outliers, which are further promoted as lesions according to their spatial probability of being in the WM, where the WM segmentation is basically derived from T1-weighted image segmentation. Lesion-TOADS [146], on the other hand, employs a sophisticated mechanism of combining information from different MR sequences (T1-weighted, T2, PD or FLAIR) in order to simultaneously segment lesions and brain structures, while distance maps from the boundaries of structures such as CSF are used to confine the segmented lesions to typical locations.

1.5.2 Longitudinal biomarkers

In contrast to cross-sectional approaches, longitudinal methods take into account two (or more) MRI scans of the same subject from different time points to calculate brain volume changes or atrophy. Typical preprocessing steps prior to longitudinal atrophy computations include [46]: extraction of the intracranial cavity mask at baseline; correction of intensity inhomogeneities; rigid registration of follow-up scans on the baseline scan; and differential bias field correction to correct for differences in intensity inhomogeneity artefacts. Longitudinal methods for brain atrophy typically try to match two MRI scans using registration techniques and directly extract small changes in brain volume from this process. Approaches include brain edge motion analysis, voxel-based statistical analysis for voxel-based morphometry, statistical parametric mapping, and local Jacobian determinant analysis after nonlinear matching between coregistered images [153, 59, 21, 151].

In what concerns lesions, many methods focus on segmenting MS lesions at a single time point, and there is not yet a single approach, according to the review of [101], that can emerge as a standard in clinical practice for the analysis of lesion evolution over time.

Although many studies have supported the correlation between the MRI-derived biomarkers and the clinical biomarkers (especially EDSS), this correlation is not very strong. This could be explained by the functional and structural redundancies in the brain that prevent this damage to be measured by clinical measurements such as EDSS. Thus MRI-derived biomarkers seem to be more sensitive than clinical biomarkers [58, 136]. Because of the low correlation, MRI-derived biomarkers are not yet standardised and still under development [77, 32, 60, 5].

1.6 Advanced MRI sequences for MS

In spite of the high sensitivity of conventional MRI derived metrics for detecting MS lesions in the white matter, they still have several limitations such as: (a) detecting the actual spatial extent of a lesion, (b) lesion characterisation (remyelinating, demyelinating, edema, necrosis etc.), (c) understanding the disease pathogenesis etc. Advanced techniques, such as diffusion tensor imaging (DTI), magnetisation transfer imaging, proton MR spectroscopy (MRS), can address these issues and may therefore help in reducing the gap between the MRI and clinical findings. We now explain the principle of each of these techniques together with their use in MS research.

Diffusion tensor imaging
Diffusion tensor imaging (DTI) measures the diffusivity of water molecules inside the tissue. This phenomenon happens at the microscopic level, well beyond the conventional MRI resolution. The two commonly used DTI measures in MS are fractional anisotropy (FA) and mean diffusivity (MD). FA measures the directional diffusivity and MD measures the overall diffusion magnitude. In case of MS, myelin sheet damage around fibres results in a decreased anisotropy of diffusion [183] and is associated with the diffuse injury and rupture of the tissue architecture [136]. This results in an increased MD and decreased FA in the normal appearing brain tissue.

Magnetisation transfer imaging
Magnetisation transfer imaging measures the transfer of magnetisation between the free mobile protons in the water pool and the protons that are attached to macromolecules. As the myelin sheet contributes to these macromolecules, the destruction of myelin sheets can also be quantified using this technique [143]. By applying the MR sequences with and without an off resonance saturation pulse, magnetisation transfer ratio (MTR) can be computed. In MS, a decrease in MTR value is associated with the demyelination [49].

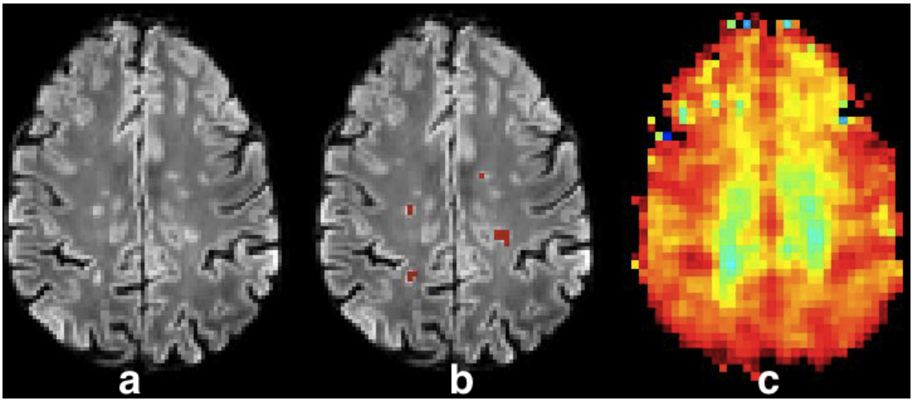


Figure 1.7: (a) Bias corrected FLAIR image followed by (b) overlaid lesion segmentation from an automated method, and (c) the acquired myo-Ins map. The myo-Ins concentration increases from red to blue. The MRSI map of the myo-Ins shows more extent of pathological changes in white matte compared to the conventional FLAIR MRI [111, 166].

Proton magnetic resonance spectroscopy
^{1}H-MRS provides information on the chemical composition of the underlying tissue. This chemical information is assessed by quantifying cellular metabolites such as N-acetyl Aspartate, choline, myo-inositol etc. Every metabolite has a function and its increase or decrease gives an indication about the health of the underlying tissue such as the state of the myelin sheets in case of MS. As chemical changes are very sensitive to the underlying pathology compared to conventional MR sequences [111, 166], MRS can play a very important role in understanding pathophysiology, diagnosis of disease in its very early stage and characterisation of lesions (see Figure 1.7). We describe MRS more in detail in the following section.

1.7 Magnetic resonance spectroscopy (MRS)

1.7.1 Metabolites in MS

In MS, the most frequently quantified metabolites are N-acetyl Aspartate (NAA), myo-inositol (myo-Ins), glutamate and glutamine (Glx), choline (Cho) and creatine (Cr). NAA is produced in the mitochondria (power house) of neurons thus it is a marker for neuronal health. In the early phase of MS, the NAA concentration increases and decreases representing the remyelination and demyelination, and as the disease progresses its concentration decreases.

Table 1.2: Evidence for the relevance of NAA metabolite in MS.

Region	**Findings**	**Study population**	**Reference**
Lesion	Reduced concentration *w.r.t.* controls	RRMS (n=9), SPMS (n=10), PPMS (n=6), Controls (n=9)	[37]
NAWM	Initial increase followed by partial recovery	2 year early follow-up of 20 RRMS patients and 10 controls	[167]
NAWM (corpus callosum)	Reduced concentration *w.r.t.* controls	11 RRMS patients and 9 controls	[79]
NAWM	Reduced concentration *w.r.t.* controls	RRMS (n=13), PPMS (n=15), Controls (n=20)	[164]
CGM	Reduced concentration in SPMS *w.r.t.* RRMS and controls	RRMS (n=15), SPMS (n=15), Controls (n=8)	[140]
DGM (thalami)	Reduced concentration *w.r.t.* controls	14 RRMS patients and 14 controls	[29, 6]
Whole brain	NAA decreased 3.6 faster than atrophy	42 RRMS patients	[62]

NAWM = normal appearing white matter, CGM = cortical grey matter, DGM = deep grey matter, RRMS = Relapsing-remitting multiple sclerosis, SPMS = Secondary progressive multiple sclerosis, PPMS = Primary progressive multiple sclerosis.

myo-Ins is a specific marker of astrocytes [141] and its increase in MS indicates the increase in astrocytes that are produced to support and repair neurons. Glx is the sum of two metabolites: glutamate and glutamine, as their separate quantification is not easy. Glx is a marker of increased glutamate-mediated excitotoxicity or increased glutamine concentration. Glutamate-mediated excitotoxicity is the pathological process in which the neurons are killed due to excessive activation by neurotransmitters such as glutamate. In a normal process, glutamate is converted into glutamine by astrocytes and then back to glutamate by the neuron. In MS, as the neurons die, glutamate synthesis is diminished, and thus, the glutamine concentration increases as the disease progresses. Cho consists of contributions from the tri-methylamine groups of glycerophosphocholine, phosphocholine, and a small amount of free choline itself [9]. Cho is found in all cell membranes and is associated with changes in membrane composition, such as membrane turnover and demyelination [70]. Cr signal represents the added contribution of creatine and phosphocreatine [141]. Cr is associated with cell metabolism and its elevation is attributed to combined effects of astrocytic demyelination in lesions and oligodendrocytic remyelination. Both Cho and Cr are found in neurons and glial cells, and were shown to be much higher in oligodendrocytes and astrocytes of cell culture extracts. Excess of both cell types probably results in the rise of Cho and Cr in the MS patients [79]. NAA is the most common metabolite studied in MS because it is a neuronal marker and the evidence for its relevance in MS is summarised in table 1.2.

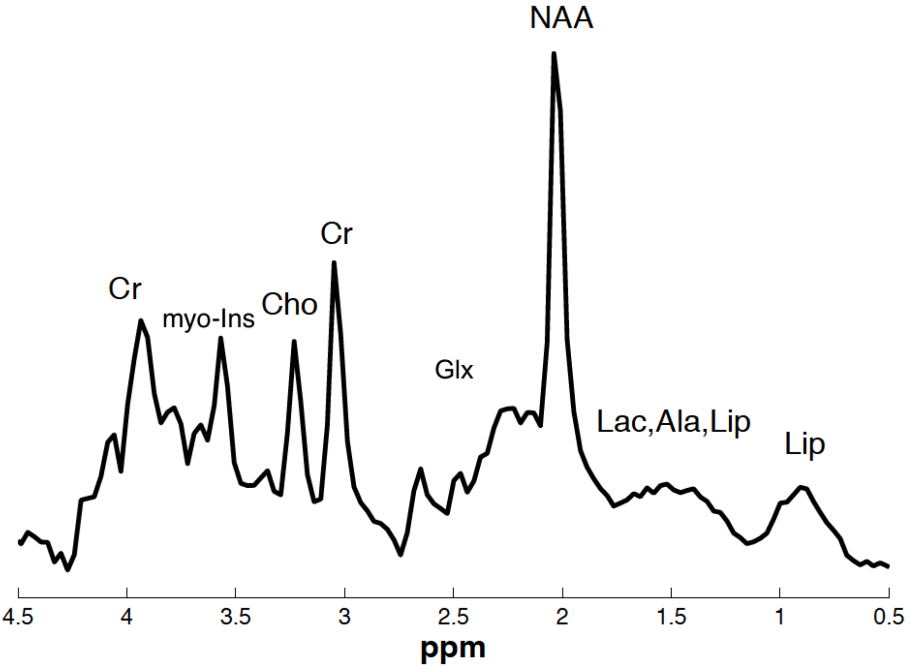


Figure 1.8: Human brain MR spectrum acquired in vivo at 1.5T showing different metabolites at their respective resonant frequencies. Echo time = 23 ms, repetition time = 6 sec, 64 averages, single voxel spectroscopy with PRESS sequence.

1.7.2 MRS acquisition

Figure 1.8 shows the healthy human brain ^{1}H-MR spectrum where the resonant frequencies of different metabolites are expressed in parts per million (ppm), and is read from right to left. NAA resonates at a frequency of 2 ppm, followed by Glx between 2.0 – 2.4 ppm, Cr with two peaks at 3.0 ppm and 3.9 ppm, Cho at 3.2 ppm and myo-Ins at 3.5 ppm. The y-axis of the spectrum represents the signal intensity, which is expressed in an arbitrary unit due to absence of a reference signal. A metabolite's concentration is related to the area under its MRS signal.

MRS spectra can be acquired using single voxel spectroscopy (SVS) or multi voxel magnetic resonance spectroscopic imaging (MRSI). In SVS, the single spectrum of a particular volume can be acquired by first suppressing the water signal followed by selecting the volume of interest using gradient pulses and finally acquiring the spectrum using *e.g.,* point-resolved spectroscopy (PRESS) or stimulated echo acquisition mode (STEAM) sequences [15]. In MRSI, a large volume is divided into multiple smaller voxels such that the spectra of every voxel can be acquired simultaneously. SVS acquisition has advantages such as higher signal to noise ratio (SNR) and better spectral quality that are counter balanced by the loss of spatial information and large voxel sizes in comparison with MRSI [139]. Both acquisition methods can be acquired at long and short echo times. Long echo time (range: 135-144 ms) results in detection of limited metabolites (*e.g.,* NAA, Cr and Cho) that have higher concentration, long T2 relaxation time and uncoupled spin systems. In case of resonance from strongly coupled spins, the signal decreases rapidly with echo time due to destructive interactions between resonances with J-coupled induced phase modulation (internal indirect interaction of two spins via the intervening electron structure of the molecule). In addition to these long echo time metabolites, short echo time (range: 20-40 ms) allows the detection of other metabolites such as myo-Ins, Glx [174].

1.7.3 Metabolite quantification

Preprocessing
Some common pre-processing steps include (a) multiplication with a filter (e.g. Lorentzian filter in SVS or Hamming filter in MRSI) in the time domain to improve the signal-to-noise ratio, (b) zero-filling for artificially improving the spectral resolution, (c) baseline correction with the aim of proper quantification of small metabolites, and (d) phasing for obtaining a pure absorption mode MR spectrum [175].

Quantification
The acquired MRSI data, after Fourier transformation from k-space to the spatial domain, can be fitted in a voxel-by-voxel fashion using a simulated metabolite basis set, matching the acquisition parameters (*e.g.,* echo time) of the given signals. The fitting can be based on time-domain MRSI signals, or after Fourier transformation to the frequency domain. The most widely used quantification software is LCModel that linearly fits the frequency domain data to a reference model using a metabolite database [126]. Another popular software package is jMRUI (Java-based magnetic resonance user interface) where the data is fitted in the time domain using prior knowledge on coupling patterns, linewidth etc [112]. Both software packages provide error estimation in data fitting using Cramer-Rao lower bound [35]. For MRSI, the output consists of a set of metabolite images ('metabolite maps') in arbitrary units, as many as metabolites in the basis set, and having the same spatial resolution as the input MRSI grid [174].

In MRSI, absolute quantification is difficult *w.r.t.* single voxel spectroscopy as obtaining the additional reference scan (water signal) is time consuming. The use of water signal as a reference scan helps in reducing several sources of errors such as differences in receiver gains *etc.* [16]. Moreover, chemical shift displacement error (spatial displacement of off-resonance signal *w.r.t.* resonance frequency) is more profound in MRSI due to simultaneous RF slice selection pulses that result in a spatial displacement of the resonances within the excitation volume.

In many studies, metabolites are measured relative to the Cr concentration, which is considered to be relatively constant in many diseases including MS. However, in some studies it has been shown that the Cr level may vary, thus, providing an error in measurements. Therefore, absolute quantification should be envisaged [87].

1.7.4 Problems with MRS imaging

Despite the fact that MRSI provides attractive complementary information with respect to conventional MRI, it is still not widely used in practice. Possible reasons include (a) low signal to noise ratio, (b) poor spatial resolution, (c) prolonged acquisition time, (d) lack of standardised protocols, (e) and limited quality control. Since the metabolite concentrations are very low, a large voxel size is required in order to capture sufficient signal, resulting in a poor spatial resolution (typically around 1 cm^3). Additionally, this results in a broad point spread function, which further causes spectral signal leakage in the neighbouring voxels. This limits the ability to differentiate among spatial features of interest, *e.g.,* between different tissue types. High spatial resolution MRSI cannot be acquired in the clinical routine with standard acquisition sequences, as it requires a long scanning time, *e.g.,* in the range of tens of minutes. Therefore, there is a need to design alternative, non-conventional approaches to create high-resolution MRSI-based metabolic images within a limited acquisition time.

1.8 Objectives and main contributions

The main objective of this thesis was to develop an automated method for combining the information from a low-resolution MRSI scan with a high-resolution MRI scan with the goal of up-sampling the spectroscopic image. Using these high-resolution or super-resolution images, this thesis aimed to better characterise small MS lesions. To this end, several data processing steps were designed and optimised: lesion segmentation in MRI, matching the image resolutions between the different modalities, multi-modal data fusion and metabolite quantification.

We now list our main contributions along with the associated publications on MS lesion segmentation and MRSI resolution enhancement:

- An automatic single time point MS lesion segmentation method that fragments the MR image of an MS brain into GM, WM, CSF and lesions. The method takes as input 3D T1-weighted and 3D FLAIR MR images and is optimised using an expectation-maximisation algorithm (chapter 2). We show good accuracy and reproducibility of our method on two clinical datasets [86, 25].
- A patch-based super-resolution method to increase the spatial resolution of a low resolution MRSI scan. The method uses high-resolution 3D T1-weighted and 3D FLAIR MR images together with brain segmentations to regularise the super-resolution process (chapter 3). Validation on synthetic and real images shows that our method preserves tissue contrast and structural information; and matches well with the trend of acquired high-resolution MRSI [85, 84].
- Two time points white matter lesion segmentation method, that jointly segments MS lesions of two time points. The method takes as input (a) 3D T1-weighted and 3D FLAIR MR images of the two time points, (b) FLAIR based subtraction image, and (c) single time point lesion segmentations of the two time points; and is optimised using a joint expectation-maximisation algorithm (chapter 4). Validation on two clinical datasets shows that our method is both accurate and consistent in segmenting MS lesions [83].
- A multi-center study on the clinical relevance of the new and the enlarging lesion volume (computed from the two time points lesion segmentation method) in RRMS (chapter 5). We observe that in majority of the patients the new and enlarging lesion volume correlates better with the clinical biomarker, EDSS score, at patient level than at a group level.

This manuscript is structured as follows.

- We first present the single time point MS lesion segmentation method in chapter 2 that forms the basis for the super-resolution and two time points lesion segmentation methods.
- Using the brain segmentation information, patch-based super-resolution of MRSI is presented in chapter 3.
- Chapter 4 presents the two time points lesion segmentation method, which takes as input single time point lesion segmentation information of the two time points.

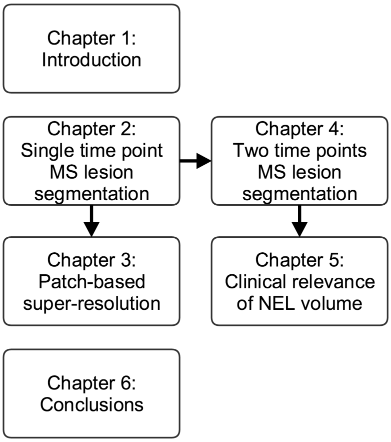


Figure 1.9: The graphical representation of the thesis. NEL = new and enlarging lesions.

- New and enlarging lesions (NEL) volume, which is computed using the two time points lesion segmentation method, is investigated in chapter 5 for its association with the EDSS score.
- Finally, we draw the conclusions and formulate some future perspectives (chapter 6).

The structure of this thesis is presented in figure 1.9.

Chapter 2

Single time point MS lesion segmentation

Abstract

The location and extent of white matter lesions on MRI are important criteria for diagnosis, follow-up and prognosis of MS. Clinical trials have shown that quantitative values, such as lesion volumes, are meaningful in MS prognosis. In this chapter, we present MS**metrix**-cross, an accurate and reliable automatic method for lesion segmentation based on MRI. In MS**metrix**-cross, 3D T1-weighted and FLAIR MR images are used in a probabilistic model to detect WM lesions as an outlier to normal brain while segmenting the brain tissue into GM, WM and CSF. The actual lesion segmentation is performed based on prior knowledge about the location (within WM) and the appearance (hyperintense on FLAIR) of lesions. We show good accuracy and reproducibility of our method in two clinical datasets which compares favourably with other publicly available MS lesion segmentation algorithms.

2.1 Introduction

The location and extent of white matter lesions on MRI are important criteria for diagnosis, follow-up and prognosis of MS. Clinical trials have shown that

---

The work presented in this chapter has been published in: Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., Daams, M., et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical 8* (2015), 367–375.

lesion volumes are meaningful outcomes for disease prognosis [170]. Accurately measuring lesion volumes is, therefore, of considerable interest in clinical practice. Manual segmentation of MS lesions is time consuming and suffers from large intra- and inter-observer variability. Therefore, in clinical trials and research studies, semi-automated methods are increasingly used. The ultimate aim, however, is to use a fully automated lesion segmentation approach, as it can further decrease the observer dependency as well as the time needed from the expert. This will be especially important for large clinical trials, as a huge amount of data needs processing. For clinical practice, an automated method would make it possible to measure lesion volumes, thereby further standardising and quantifying the MRI reading.

In this chapter, MS**metrix**-cross is introduced, a robust automatic method for WM lesion segmentation based on 3D T1-weighted and 3D FLAIR MR images. The method is independent of scanner and acquisition protocol and does not require a training image database of expert lesion segmentations. The proposed method models the WM lesions as an outlier class while classifying the brain into three classes, GM, WM and CSF using a healthy brain atlas. The aim of this work is to validate the proposed lesion segmentation method on two distinct datasets from two clinical centres, used to assess accuracy as well as reproducibility. In addition, results are compared both quantitatively and qualitatively with two well-known and publicly available automatic unsupervised lesion segmentation software implementations: LST [142] and Lesion-TOADS [146].

2.2 Methods

This section provides more details on the proposed lesion segmentation method and summarises the publicly available lesion segmentation methods we compare to. Furthermore, the data and the validation tests used for evaluation of the methods are described.

2.2.1 Method description

Figure 2.1 shows a schematic overview of the proposed method, which takes as inputs 3D T1-weighted and a 3D FLAIR image acquired from an MS patient. The images are preprocessed and co-registered before executing the main loop of the algorithm, consisting of brain tissue segmentation, lesion segmentation and lesion filling. We describe each of the steps into more detail below.

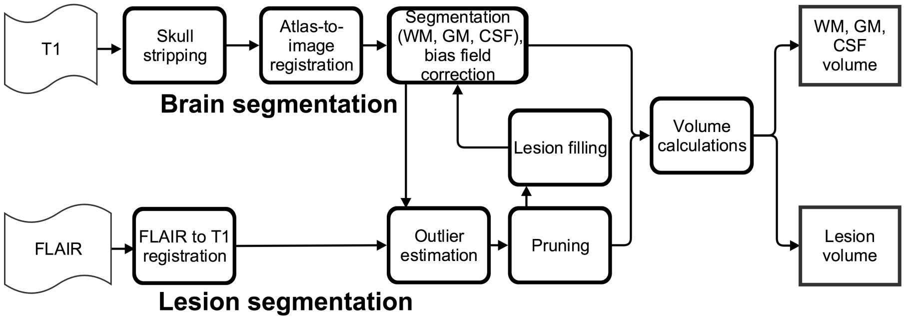


Figure 2.1: Schematic representation of the MS**metrix**-cross method.

The first step, the preprocessing, has three stages:

- In the first stage, the input FLAIR image of the patient is rigidly co-registered with the input T1-weighted image using the normalised cross correlation coefficient as a similarity measure.
- In the second stage, the T1-weighted input image is skull stripped classifying each voxel either as a brain region or a non-brain region based on the affine registration of a brain mask available from the MNI-atlas using the normalised cross correlation coefficient as a similarity measure followed by a non-rigid registration using the normalised mutual information as a similarity measure.
- In the third stage, the probabilistic anatomical priors for GM, WM and CSF, which are also available from the MNI-atlas, are transferred to the skull stripped T1-weighted image space using the affine transformation and non-rigid deformation computed above.

In these preprocessing stages, the rigid and affine registrations use a multi-resolution approach with a Trimmed Least Square scheme and block-matching (NiftyReg[1], [116]). The non-rigid registration is based on the Free-Form B-spline deformation model and also uses a multi-resolution approach (NiftyReg, [107]).

In the second step, which builds further on the work of [177], a probabilistic model is formulated to segment the skull stripped T1-weighted image based on prior knowledge given by the probabilistic tissue priors mentioned above. The model assumes a Gaussian distribution of the image intensities per tissue class, a smoothly varying bias field for the intensity non-uniformities and contains

---

[1] `http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg`

a spatial consistency model based on a Markov Random Field (MRF). This model is optimised using an EM algorithm [177] as implemented in NiftySeg[2], [26]. The algorithm iteratively estimates the parameters of each tissue class, as well as the bias field parameters, and maintains the spatial consistency until convergence. After the convergence of the EM algorithm, the T1-weighted image is bias field corrected and segmented into the three tissue classes, *i.e.,* GM, WM and CSF.

In the third step, an outlier class is estimated from the co-registered FLAIR image of the same patient using the three tissue class segmentations from the previous step as prior information. This is performed using the same EM algorithm as described in the second step, but now an outlier map is included [176]. Using the tissue segmentations of the T1-weighted image as prior information, the intensities of each tissue class in the FLAIR image are modelled as a normal distribution and the deviation of each FLAIR voxel from this model is estimated as an outlier belief map. This map is iteratively updated by the EM algorithm and, after convergence, the FLAIR image is bias field corrected and an outlier belief image is produced. This outlier belief image is used as an initialisation for the MS lesion segmentation in the next step.

In the fourth step, we segment the lesions in the outlier map, *i.e.,* we 'prune' the outlier map, as not every outlier is a lesion (*e.g.,* the outlier map might include partial volume effects, artefacts, etc.). In order to differentiate the MS lesions from such non-lesion outliers, some extra a priori information about the location and the appearance of the lesions needs to be incorporated. The outliers need to be in the WM region, where the WM segmentation is derived from T1-weighted image segmentation in the second step. Moreover, the underlying intensities of the outliers should be hyperintense compared to the GM intensities on bias field corrected FLAIR image. The hyperintensity is defined as a threshold that is equal to the mean plus two times the standard deviation of GM intensities in the bias field corrected FLAIR image. Furthermore, another mask based on co-registered T1-weighted image atlas, is used to exclude areas (WM in-between the ventricles and in the cortical regions near the great longitudinal fissure) that are prone to show false lesion segmentation. Finally, each masked outlier voxel needs to have a minimum number of 5 adjacent outlier voxels (empirically determined), in order to avoid spurious lesion detection.

In the fifth step, this lesion segmentation is then used to fill in the lesions in the bias corrected T1-weighted image with WM intensities. This global lesion filling method is very similar to the LEAP method [28] where the mean and standard deviation of WM intensities are computed using the WM segmentation from the second step. In our case, the WM segmentation is a soft segmentation and

---

[2] `http://niftyseg.sourceforge.net/index.html`

therefore, is thresholded to values above 0.5, in order to make sure that only WM intensities are included in the statistics. Finally, the whole filled-in volume is smoothed with a Gaussian kernel of radius 1 voxel, and normalisation is applied in order to restore the global standard deviation of the WM intensities as in [28].

Steps 2-5 are repeated until there is no significant change in the tissue and the lesion segmentation. The idea of repeating the second and the third step is that the lesions are primarily WM, therefore, the T1-weighted lesion filling will result in better brain tissues segmentation, which in turn results in better segmentation of lesions due to the improved WM mask used in the fourth step. After the last iteration lesions are recovered from the GM segmentation in case the outlier belief is high. This is because some WM lesions that are very close to cortical GM might be wrongly segmented as GM and therefore, are missed in the pruning step, *i.e.,* the fourth step. These lesions are added to the previously found lesions in the fourth step. Subsequently the T1-weighted image is filled one more time and segmented again providing the final segmentations of WM, GM and CSF.

2.2.2 Comparison with state-of-the-art methods

For comparison, the validation tests performed for MS**metrix**-cross[3] were executed on the same datasets using two state-of-the-art software packages: LST and Lesion-TOADS. Unless otherwise stated, default parameter settings were used for all implementations, thus parameter tuning was not performed neither at dataset level, nor at patient level.

LST

LST (Lesion Segmentation Tool) v1.2.3 is implemented in SPM8[4] and is based on a lesion growth algorithm described in [142] . In short, LST determines GM, WM and CSF segmentations from T1-weighted images and computes the FLAIR intensity distributions of these tissue classes. The amount of 'hyperintensity' of each voxel in terms of distance from the mean intensity of the WM, GM and CSF distributions in the FLAIR image is crucial for defining a conservative lesion belief map (obtained by thresholding the GM belief map) and a liberal lesion belief map (consisting of the sum of the three lesion belief maps). Lesion

---

[3] MS**metrix**-cross pipeline presented here is the version 1.2 of the clinical product – MS**metrix**.

[4] `http://www.fil.ion.ucl.ac.uk/spm/software/spm8/`

growing is then performed iteratively between the conservative and the liberal belief maps, until no more voxels are added to the lesions.

Lesion-TOADS

Lesion-TOADS [146] is available as a plug-in for the MIPAV software[5]. Lesion-TOADS implements an iterative algorithm for fuzzy classification of the image intensities, using a combination of topological and statistical atlases. An additional lesion class is added to the brain segmentation model, using the same spatial prior as WM; lesions and WM are then separated by selecting, inside the grouped region, whichever has the higher membership value. Prior knowledge about areas where false positives commonly occur is used to define penalty weights based on the distance to these areas (*e.g.,* distance to ventricles, GM structures and interventricular WM). This method segments multichannel input images simultaneously, using an intensity-weighting scheme that optimises the effect of each channel onto the segmentation of each tissue class. Although it is not a default setting, the method is also able to estimate bias field corrections to deal with local intensity non-uniformities. For a fair comparison against MS**metrix**-cross and LST, where bias field corrections are included by default, this option was also turned on for Lesion-TOADS, since otherwise its results were not competitive.

2.2.3 Data

Dataset 1

20 MS patients participated in a study at VU University Medical Center, Amsterdam, the Netherlands. The study was approved by the local ethics committee and all patients signed informed consent forms. MR imaging was performed on a 3T whole body scanner (GE Signa HDxt, Milwaukee, WI, USA). The protocol contained two 3D sequences: a fat-saturated 3D FLAIR (TR: 8000 ms, TE: 125 ms, TI: 2350 ms, 250 × 250 mm^2 field of view (FOV), 132 sagittal slices, 0.98 × 0.98 × 1.2 mm^3 voxel resolution) and a 3D T1-weighted fast spoiled gradient echo (FSPGR) sequence (TR 7.8 ms, TE 3 ms, FA 12°, 240 × 240 mm^2 FOV, 176 sagittal slices, 0.94 × 0.94 × 1 mm^3 voxel resolution). Reference WM lesion segmentations were constructed manually using the 3D FLAIR and 3D T1-weighted images by a highly trained neuroradiological team (Steenwijk et al., 2013). In short, both 3D T1-weighted and 3D FLAIR images were co-registered and orthogonally reformatted to the axial plane, and the

---

[5] `http://mipav.cit.nih.gov/`

axially reformatted images were then used to identify and outline the lesions. Lesion identification was performed by three raters in consensus using the 3D FLAIR images; the raters were allowed to view the corresponding co-registered 3D T1-weighted image. Lesions were only identified if they were larger than 3 voxels in-plane and visible on at least two consecutive slices. In the next step, two trained technicians manually outlined the identified lesions on the 3D FLAIR using MIPAV. The expert lesion segmentation resulted in a wide range of lesion volumes going from 1.88 to 50.95 ml, with a mean of 16.33 ml and a standard deviation of 11.49 ml. For comparison purpose, these lesion segmentations were resampled from the FLAIR image space to their corresponding T1-weighted image space using nearest neighbour interpolation as implemented in NiftyReg.

Dataset 2

Dataset 2 contains scans from 10 MS patients scanned twice, with re-positioning (time interval between two scans is 5 $\sim$10 minutes), on each of three different 3T scanners from GE (Discovery MR750w), SIEMENS (Skyra) and PHILIPS (Achieva). The protocol contained two 3D sequences: T1-weighted and FLAIR, and their details are described in table 2.1. For this dataset, no expert segmentations were available. Also, the resolution of T1-weighted and FLAIR images from all the scanners is high and therefore, due to very high computational memory requirement, none of the methods were able to run on these high resolution images; therefore, the T1-weighted image was down sampled to ($1 \times 1 \times 1$ mm^3) resolution. The FLAIR image was not down sampled at this point because it is rigidly co-registered to T1-weighted image in the initial stage of the method and thus will have the T1-weighted image resolution. This study was carried out in accordance with the recommendations of the ‘International Conference on Harmonisation of Good Clinical Practice (ICH-GCP)’, and the applicable Belgian and Dutch legislation. The study was approved by the UZ Brussels ethical committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

2.2.4 Performance tests

Comparison to expert segmentations on dataset 1
For dataset 1, the agreement between automatic methods and expert reference segmentation is evaluated at a voxel-by-voxel level. Spatial agreement is reported by the Dice similarity index [43]), defined as the ratio between the number of voxels where both the automatic and the expert reference segmentation agree (true positives) and the mean number of voxels labelled as lesion by the two methods. Additionally for dataset 1, the average Dice performance for each

Table 2.1: Dataset 2 sequences description for all three scanners.

Sequence	TR (ms)	TE (ms)	TI (ms)	FOV (mm^2)	Voxel size (mm^3)	No. of slices (sagittal)
GE						
3D T1-weighted FSPGR	7.32	3.14	NA	220 × 220	0.43 × 0.43 × 0.50	328
Fat saturated 3D FLAIR	9500	135.78	2428	240 × 240	0.47 × 0.47 × 0.70	232
SIEMENS						
3D T1-weighted MPRAGE	2300	2.29	NA	240 × 240	0.94 × 0.94 × 0.94	176
Fat saturated 3D FLAIR	5000	387	1800	230 × 230	0.45 × 0.45 × 0.90	192
PHILIPS						
3D T1-weighted FSPGR	4.93	2.3	NA	230 × 230	0.53 × 0.53 × 0.50	310
Fat saturated 3D FLAIR	4800	276	1650	240 × 240	1.04 × 1.04 × 0.56	321

NA = Not available.

method is computed separately for patients with small, medium and large lesion volumes in order to assess whether the methods' performance is depending on lesion volume. For the total lesion volume, agreement between automatic and expert reference segmentation is evaluated through the intra-class correlation coefficient (ICC) and the absolute volume difference. ICC is a measure assessing the agreement of measurements made by multiple observers measuring the same quantity [93]. In this chapter, ICC is used in the absolute agreement formulation. The absolute volume difference is computed as the absolute difference between the total volume reported by the automatic method and the corresponding value derived from expert reference segmentation.

For dataset 1, the automatic methods segmentation quality is evaluated at a voxel-by-voxel level. The ability to segment lesions is reported by sensitivity and is defined as the ratio between true positives and the total number of lesion voxels in the expert reference segmentation (true positives and false negatives). The relevance of the segmentation is measured by precision and is defined as the ratio between true positives and the total number of lesion voxels in the automatic segmentation (true positives and false positives). Since MS**metrix**-cross is an iterative method, the benefit gained over the iterations is verified by reporting the Dice similarity index for dataset 1.

To determine if there is a statistical difference between MS**metrix**-cross and LST and between MS**metrix**-cross and Lesion-TOADS methods' performance, a two tailed paired t-test is performed, except for the average Dice performance as a function of lesion volume, where the sample size is small.

Lesion volume reproducibility assessment on dataset 2

For dataset 2, ten patients were examined twice in each scanner. For the total lesion volume, agreement between the first and the second scan for each scanner is evaluated through Dice similarity index and absolute volume difference. Here, a two tailed paired Wilcoxon signed-rank test is performed instead of a t-test as the Dice similarity index and absolute volume difference for all methods are not normally distributed.

2.3 Results

2.3.1 Accuracy results on dataset 1

In this section we first present quantitative results (Dice similarity index, ICC and absolute volume difference) followed by the qualitative results, where the segmentation results are evaluated visually by presenting the best and the worst cases for each of the automatic methods.

Quantitative results

Table 2.2 presents the quantitative results where the Dice similarity index reaches an average of 0.67 for MS**metrix**-cross, followed by 0.61 for Lesion-TOADS and 0.55 for LST. In order to visualise the volumetric correlation of each automatic method to the expert reference segmentation, Figure 2.2 shows the scatter plots for total lesion volume (lesion load) in ml of each method compared to the volume of expert reference segmentation. MS**metrix**-cross is well correlated to the expert reference lesion volume, but has a general trend of slightly underestimating the lesion volume compared to the expert reference lesion segmentation. Lesion-TOADS shows an even stronger trend of volume underestimation. The absolute volume error and total lesion volume are better correlated to the expert reference values for LST (average 4.75 ml absolute volume error with an ICC of 0.87) than for MS**metrix**-cross and Lesion-TOADS (5.15 ml, 0.80, and 6.8 ml, 0.63 respectively).

In order to provide a better understanding of the segmentation performance, also the sensitivity and precision at voxel level are presented in Table 2.3 for each of the automatic methods compared to the expert reference segmentation. The sensitivity reaches an average of 0.57 for MS**metrix**-cross, followed by 0.50 for Lesion-TOADS and for LST. MS**metrix**-cross is more precise, with an average precision of 0.83 compared to 0.70 and 0.81 for LST and Lesion-TOADS, respectively. Table 2.4 presents the average Dice performance of each method for patients with small, medium and large lesion volumes, computed according to the expert reference segmentation. For this, the dataset was divided into three intervals according to lesion load: $\leq$5 ml, 5–15 ml and $\geq$15 ml. From the table it can be concluded that all methods show an increase in their performance for

Table 2.2: Agreement measures (Dice similarity index, ICC and absolute volume difference) between automatic and expert reference lesion segmentation for MS**metrix**-cross, LST and Lesion-TOADS for 20 MS patients.

Automatic method	**Dice**	**ICC**	**Absolute volume difference (ml)**
MS**metrix**-cross	0.67 ± 0.11	0.8	5.15 ± 4.75
LST	0.55 ± 0.16**	0.87	4.75 ± 3.63
Lesion-TOADS	0.61 ± 0.09**	0.63	6.86 ± 5.70**

Dice and absolute volume difference are presented in mean ± standard deviation.
** Values significantly different fromMS**metrix**-cross (paired t-test with $p < 0.01$ significance level).

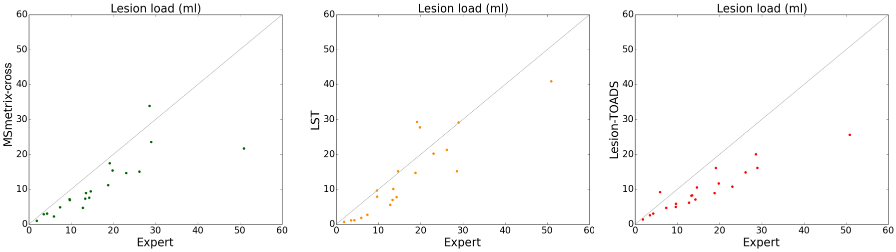


Figure 2.2: Scatter plots of expert reference values versus automatically computed values for total lesion volume (ml). The three columns show results for MS**metrix**-cross, LST and Lesion-TOADS, respectively.

large lesion volume, with MS**metrix**-cross being most consistent, *i.e.*, having a higher stable range of values among the groups.

The Dice similarity index for MS**metrix**-cross grows from 0.51 ± 0.14 at

Table 2.3: Segmentation quality measures (sensitivity and precision) between automatic and expert reference lesion segmentation for MS**metrix**-cross, LST and Lesion-TOADS for 20 MS patients.

Automatic method	**Sensitivity**	**Precision**
MS**metrix**-cross	0.57 ± 0.13	0.83 ± 0.11
LST	0.50 ± 0.22*	0.70 ± 0.09**
Lesion-TOADS	0.50 ± 0.08**	0.81 ± 0.17

Sensitivity and precision are presented in mean ± standard deviation.
* Values significantly different from MS**metrix**-cross (paired t-test with $p < 0.05$ significance level).
** Values significantly different from MS**metrix**-cross (paired t-test with $p < 0.01$ significance level).

Table 2.4: Agreement measure (average Dice similarity index) for small (n=3), medium (n=9) and large (n=8) lesion volumes for automatic methods. Here, the t-test is not performed, as the sample size is small for each group.

Automatic method	**Average Dice**		
	(< 5) ml	(5-15) ml	(>15) ml
MS**metrix**-cross	0.61	0.62	0.74
LST	0.33	0.51	0.69
Lesion-TOADS	0.52	0.58	0.67

iteration 1, to 0.60 ± 0.13 at iteration 2, to 0.66 ± 0.11 at iteration 3, which proves the benefit of using the proposed iterative approach (see figure 2.1). Stopping after iteration 3 provides a reasonable trade-off between sensitivity and precision.

Qualitative results

In the following figures of this section, the original FLAIR image is shown followed by the lesion segmentation from the expert reference segmentation (yellow), MS**metrix**-cross (green), LST (orange) and lesion-TOADS (red) are super-imposed on the bias corrected FLAIR images.

Figure 2.3 shows the best case for MS**metrix**-cross (Dice: 0.84, sensitivity: 0.84 and precision: 0.83),which is also the best case for Lesion-TOADS (Dice: 0.79, sensitivity: 0.73 and precision: 0.87). This case has Dice: 0.69, sensitivity: 0.87, and precision: 0.57 for LST. The high sensitivity and low precision of LST are caused by the presence of false positive lesions and the overestimation of lesion boundaries (marked by cyan arrow heads) compared to the other two methods. Between MS**metrix**-cross and Lesion-TOADS, the higher Dice similarity index of MS**metrix**-cross is because of the higher sensitivity. Notice that the lesions marked by pink arrow heads were picked up by MS**metrix**-cross but not by the other methods except one by Lesion-TOADS.

Figure 2.4 shows the worst case for MS**metrix**-cross (Dice: 0.45, sensitivity: 0.31 and precision: 0.79). LST and Lesion-TOADS had comparable performance, namely, Dice: 0.43, sensitivity: 0.31, and precision: 0.70 for LST, and Dice: 0.47, sensitivity: 0.35, and precision: 0.73 for Lesion-TOADS. Some subtle lesions are either missed or the lesions are underestimated (purple arrowhead). This accounts for the low sensitivity and the low Dice similarity index for all the methods. More examples from dataset 1 are given in the appendix A.

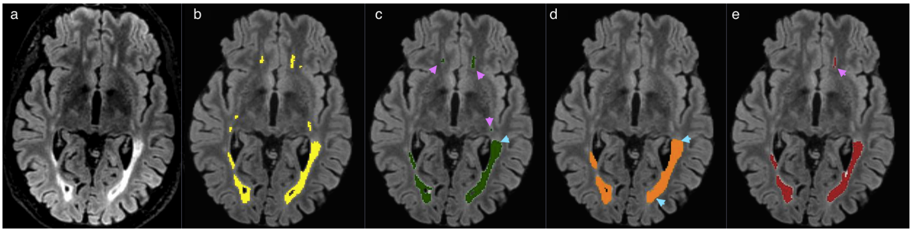


Figure 2.3: Original FLAIR image (a) followed by bias corrected FLAIR image and superimposed lesion segmentation from: (b) expert segmentation, (c) MS**metrix**-cross, (d) LST, and (e) Lesion-TOADS. Cyan arrowheads show false positive lesions and overestimation of the lesion boundaries in LST. Pink arrow heads show lesions picked by MS**metrix**-cross but not by the other methods except one in Lesion-TOADS.

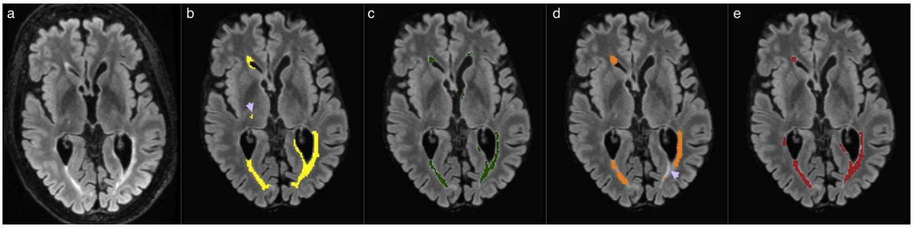


Figure 2.4: Original FLAIR image (a) followed by bias corrected FLAIR image and superimposed lesion segmentation from: (b) expert segmentation, (c) MS**metrix**-cross, (d) LST, and (e) Lesion-TOADS. Purple arrow heads show some subtle lesions that are either missed or the lesions are underestimated.

2.3.2 Reproducibility results on dataset 2

In this section, we first present quantitative results (Dice similarity index and absolute volume difference) followed by the qualitative results, where visual results are presented for MS**metrix**-cross, LST and Lesion-TOADS.

Quantitative results

Table 2.5 presents the quantitative results where the Dice similarity index reaches an average of 0.71 for LST, followed by 0.69 forMS**metrix**-cross and 0.63 for Lesion-TOADS. The absolute lesion volume difference is less for LST (average 0.44ml absolute volume difference) than for MS**metrix**-cross and Lesion-TOADS (0.54 and 1.58 ml respectively).

In order to visualise the volumetric agreement between scan 1 and scan 2 of the corresponding automatic methods, Figure 2.5 contains the Bland-Altman plot

Table 2.5: Agreement measures (Dice similarity index and absolute volume difference) between scan 1 and scan 2 of the corresponding automatic methods.

Automatic method	**Dice**	**Absolute volume difference (ml)**
MS**metrix**-cross	0.69 ± 0.14	0.54 ± 0.58
LST	0.71 ± 0.18	0.44 ± 0.69
Lesion-TOADS	0.63 ± 0.17**	1.58 ± 2.2

Dice and absolute volume difference are presented in mean ± standard deviation.
** Values significantly different from MS**metrix**-cross (paired t-test with $p < 0.01$ significance level).

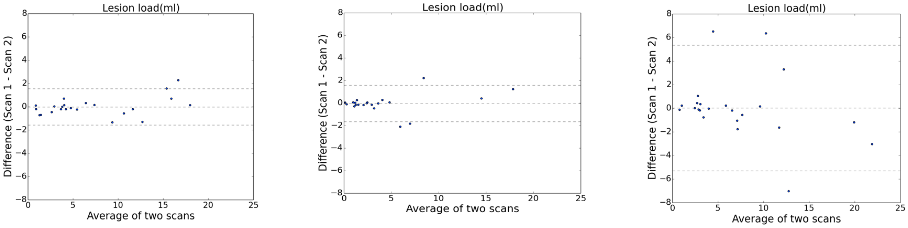


Figure 2.5: Bland-Altman plots for total lesion volume agreement between scan 1 and scan 2 of the corresponding automatic methods. The three columns show results for MS**metrix**-cross, LST and Lesion-TOADS, respectively.

for total lesion volume of each method at scan 1 compared to the volume at scan 2.

Qualitative results

In the following figures of this section, the original FLAIR image is shown followed by the lesion segmentations from MS**metrix**-cross (green), LST (orange) and Lesion-TOADS (red) are super-imposed on the bias corrected FLAIR images. In each figure, the first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2.

Figure 2.6 shows the best case (highest reproducibility) for MS**metrix**-cross (Dice: 0.84). In this case, LST has the same performance (Dice: 0.84) as MS**metrix**-cross followed by Lesion-TOADS (Dice: 0.65). A good Dice similarity index for both MS**metrix**-cross and LST is mainly due to the fact that both methods are consistent in the lesion segmentation in their respective scan 1 and scan 2, while Lesion-TOADS seems to differ more in its lesion

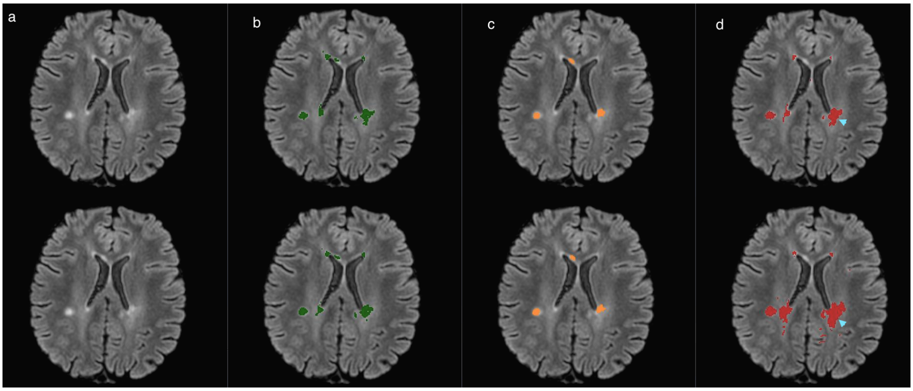


Figure 2.6: Bias corrected FLAIR image (a) and super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) LST, and (d) Lesion-TOADS. The first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2. Cyan arrow heads show the difference in the lesion segmentation boundary between scan 1 and scan 2 for Lesion-TOADS.

segmentation boundary between scan 1 and scan 2 (marked by cyan arrow head) and thus resulting in a lower Dice similarity index.

Figure 2.7 shows the worst case for MS**metrix**-cross (Dice: 0.38). In this case, Lesion-TOADS has the best performance (Dice: 0.52) followed by LST (Dice: 0.46). The low Dice similarity index for MS**metrix**-cross is caused by false lesion detection in either of the scans (marked by cyan arrow heads). The higher Dice similarity index for Lesion-TOADS compared to MS**metrix**-cross and LST is mainly due to its quite consistent performance in detecting both true and false lesions in scan 1 and scan 2 (marked by pink and cyan arrow heads) for this case. For LST, a lower Dice similarity index is mainly due to the fact that it detects several few false lesions in either of the scans (marked by cyan arrow heads). However, both MS**metrix**-cross and Lesion-TOADS are more sensitive and thus detect subtle lesions (marked by pink arrow heads) whereas LST misses them in most of the cases (marked by purple arrow head). More examples from dataset 2 are given in the appendix A.

2.4 Discussion and conclusions

In this chapter, a robust method for WM lesion segmentation is proposed, incorporating a lesion-filling algorithm for T1-weighted images and an iterative process of using the lesion-filled T1-weighted image for WM/GM/CSF tissue

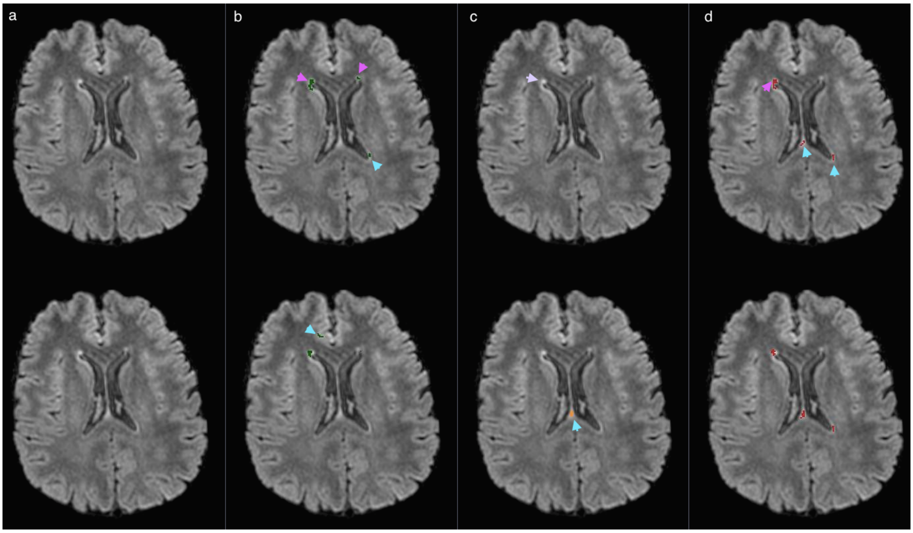


Figure 2.7: Bias corrected FLAIR image (a) and super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) LST, and (d) Lesion-TOADS. The first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2. Cyan arrowheads show the difference in the lesion segmentation between scan 1 and scan2 for MS**metrix**-cross, LST and Lesion-TOADS. Pink arrow heads show subtle lesions that are picked up by MS**metrix**-cross and Lesion-TOADS. Purple arrow head shows missed subtle lesions by LST.

segmentation and the FLAIR image for lesion detection and segmentation. The method is fully automatic and has proven to be robust for different scanners without parameter tuning. As opposed to previous work [176, 146], where multi-channel images were used simultaneously for lesion segmentation, we adopt an approach that tries to imitate more the human expert. In our case, T1-weighted and FLAIR images are used independently in order to fully exploit the main characteristics of each modality. In this chapter, the method is evaluated for 3D T1-weighted and 3D FLAIR MR images. However, first visual assessment of the methods' performance with 3D T1-weighted and 2D FLAIR MR images indicates good results.

The problem of MS lesion segmentation has received interest in the research community for the past 15 years (see [61, 108] for recent reviews). Although WM lesions are visible on the FLAIR image, automatic detection and delineation remain very challenging. The differentiation of MS lesions from 'dirty' WM and CSF pulsation artefacts in FLAIR is very difficult, as they share typical spatial locations and appear similar to MS lesions. Moreover, the MS lesions that are visible on FLAIR images exhibit very high pixel intensity variations, thus lesions can be classified as hyper- or hypo-intense. While it is quite easy to identify hyper-intense lesions, for example by intensity thresholding, hypo-intense lesions, on the other hand, are iso-intense with GM intensities and thus pose a challenge.

Manual lesion segmentation is time-consuming and suffers from inter- and intra-rater variability, with studies reporting low variability of total lesion volume estimation only in ideal situations, *e.g.,* when lesion detection is done by a single observer and lesion delineation is performed by multiple raters, aided by the same semi-automated software tool ([52]: coefficient of variation of 2.3% for intra-observer and 2.9% for inter-observer). However, such good results are not reproduced in the realistic situation when lesion detection and delineation are done by experts in different centres; as an example, MR images from 24 MS patients have been segmented by 2 experts for the MICCAI 2008 challenge[6] and variability reached a relative volume difference of 68% and a percentage of false negatives (number of lesions marked by one rater only) of 32%, as reported by [61]; section 5.3. Semi-automatic lesion segmentation could mean that lesion detection is done by the expert user and an automatic method for 'lesion growing' (*i.e.,* lesion segmentation) is applied afterwards. This approach is more consistent than manual delineation, but still time consuming and not reproducible.

Automatic methods have the obvious advantage of being consistent and fast when compared to manual or semi-automatic methods [179]. These automatic methods can broadly be classified into supervised and unsupervised category. As previously described in the introduction section (1.5.1), supervised methods require a training dataset and perform well only when the new image to be segmented is well represented in the training dataset. On the other hand, unsupervised methods such as our proposed method, LST and Lesion-TOADS do not require training images. They perform brain segmentation into GM, WM and CSF (with or without lesion detection) and often rely on post-processing approaches in order to segment lesions (*e.g.,* lesion growing or pruning). The assumptions that are made in order to segment lesions have a great impact on the results. From this point of view, MS**metrix**-cross has similarities to LST, since both methods detect FLAIR-hyper-intense outliers, which are further promoted as lesions according to their spatial probability of being in the WM, where the WM segmentation is basically derived from T1-weighted image segmentation. Nevertheless, MS**metrix**-cross's iterative process is meant to allow a more reliable estimation of the WM mask, as the WM/GM/CSF segmentation of the lesion-filled T1-weighted image should improve at each iteration. Lesion-TOADS, on the other hand, employs a sophisticated mechanism of combining information from different MR sequences (T1-weighted, T2, PD or FLAIR) in

---

[6] `http://www.ia.unc.edu/MSseg/`

order to simultaneously segment lesions and brain structures, while distance maps from the boundaries of structures such as CSF are used to confine the segmented lesions to typical locations. These mechanisms are bound to have a large number of parameters that require tuning. The effect of keeping all parameters to their default values was obvious when comparing results of Lesion-TOADS on dataset 1 and 2. Dataset 2 images were not as good as those from dataset 1 (based on visual quality control of input images), which translated into a significant decrease in segmentation performance of Lesion-TOADS in dataset 2. Based on visual quality control of the resulting segmentations, Lesion-TOADS introduced a large amount of false positives, which was not the case in dataset 1.

An important contribution of this work is the experimental validation of MS**metrix**-cross, both for accuracy and reproducibility. On dataset 1, which was used for the evaluation of the segmentation accuracy, we could observe that MS**metrix**-cross has a reasonably good overlap with the expert segmentation, but with some systematic undersegmentation. The fair sensitivity of MS**metrix**-cross suggests that it can segment subtle lesions reasonably well but still misses small lesions (lesions that are less than five voxels or lesions appearing on a single slice) or lesions at the border between GM and WM due to imperfect tissue segmentation. However, MS**metrix**-cross is quite consistent in segmenting small, medium and large lesions. False positive lesions are mostly found near the frontal horn, in the posterior end of the corpus callosum, also known as splenium, and in the cortical region near the great longitudinal fissure. On dataset 2, the reproducibility of the segmentation as well as the lesion volumes were evaluated, as the reproducibility on the test-retest data is a very important aspect to check the consistency and reliability of the results. The main sources of inconsistency are differences in the estimation of the lesion boundaries between both scans and differences in detecting smaller (often false) lesions (see *e.g.,* Figure A.4 in the appendix A).

The comparison with the existing lesion segmentation tools, LST and Lesion-TOADS, shows the outperformance of MS**metrix**-cross regarding the accuracy of the lesion segmentation. The Dice similarity index with respect to the expert reference segmentation is significantly higher for MS**metrix**-cross than for LST ($p<0.01$) and Lesion-TOADS ($p<0.01$). Looking in more details, this could mainly be explained by the significantly higher precision compared to LST ($p<0.01$) and the significantly higher sensitivity compared to Lesion-TOADS ($p<0.01$) as well as LST ($p<0.05$). LST, however, has a slightly lower average absolute volume error, although not significantly. This suggests that it misses lesions or hypo-intense lesions (see *e.g.,* Figure 2.4) and compensates for these missed lesions' volume by overestimating other, mostly larger, lesions (see *e.g.,* Figure 2.3). This also explains the increase in segmentation overlap performance of LST as a function of the lesion load. Additionally, LST often introduces false lesions in the corpus callosum (see *e.g.,* Figure A.1 in the appendix A). For Lesion-TOADS, the comparable precision but a lower sensitivity compared to MS**metrix**-cross, implies a larger amount of false positives within the Lesion-TOADS segmentation. It suffers from false lesions segmentation, which is, unlike MS**metrix**-cross, randomly distributed in the image (see *e.g.,* Figure A.2 in the appendix A) and is less accurate in segmenting the lesion boundary when compared to MS**metrix**-cross. Lesion-TOADS, like LST and MS**metrix**-cross, also shows an increase in the segmentation overlap with expert segmentation as a function of lesion load.

The first dataset presented in this chapter, including expert reference segmentations, was also used in [160], which means that it is possible to directly compare results herein to those published in [160]. The results in [160] may in fact be regarded as a best-case scenario, because the k-nearest neighbour classification employed therein is a supervised classification technique that was trained using a leave-one-out experiment on the 20 patients of dataset 1, based on the available expert reference segmentation. An optimal configuration has been sought and post-processing has been applied to reduce the number of small false positive regions. The average Dice similarity index reached a maximum of 0.75 ± 0.08 after optimal post-processing. Volumetric correspondence in terms of ICC was 0.93 after post-processing. These numbers can be compared to the MS**metrix**-cross, LST and Lesion-TOADS results in Table 2.2, keeping in mind that these methods are unsupervised and did not require any tuning or post-processing on this particular dataset.

Regarding the reproducibility of the lesion segmentation, we found a similar overlap as well as a similar absolute lesion volume difference between MS**metrix**-cross and LST. Both automatic methods have almost identical limits of agreement (see *e.g.,* Figure 2.5). In contrast, Lesion-TOADS shows much larger limits of agreement indicating clearly lower lesion volume reproducibility. Similar to MS**metrix**-cross, Lesion-TOADS is not very consistent in introducing false positive lesions in two consecutive scans (see *e.g.,* Figure 2.6), and, secondly, it is not as sensitive as MS**metrix**-cross in segmenting the lesion boundaries in the test-retest scans (see *e.g.,* Figure 2.6).

In conclusion, the proposed automatic lesion segmentation method, MS**metrix**-cross, has been demonstrated to provide more accurate lesion segmentations with a similar reproducibility compared to the state-of-the-art software tools. We believe that, through its robustness and automation, MS**metrix**-cross could bring an added value (possibility to measure lesion volumes) for the clinical routine evaluation of MS patients.

Chapter 3

Patch-based super-resolution of MRSI: Application to MS

Abstract

MRSI provides complementary information to conventional magnetic resonance imaging. Acquiring high-resolution MRSI is time consuming and requires complex reconstruction techniques. In this chapter, a patch-based super-resolution method is presented to increase the spatial resolution of metabolite maps computed from MRSI. The proposed method uses high-resolution anatomical MR images (T1-weighted and FLAIR) to regularise the super-resolution process. The accuracy of the method is validated against conventional interpolation techniques using a phantom, as well as simulated and in vivo acquired human brain images of multiple sclerosis subjects. We showed that the method preserves tissue contrast and structural information, and matches well with the trend of acquired high-resolution MRSI. These results suggest that the method has potential for clinically relevant neuroimaging applications.

The work presented in this chapter has been published in: Jain, S., Sima, D.M., Sanaei Nezhad, F., Hangel, G., Bogner, W., Williams, S., Van Huffel, S., Maes, F. and Smeets, D. Patch-based super-resolution of MR spectroscopic images: Application to Multiple Sclerosis *Frontiers in Neuroscience 11: 13* (2017), 12 pages.

3.1 Introduction

MRSI of the brain provides information on the chemical composition of tissues and may reveal underlying metabolic changes that are not visible on conventional MRI such as T1-weighted MRI. For example in MS, MRSI can show the extent of damage in normal appearing white matter and therefore is more sensitive in detecting pathological changes compared to conventional MRI [50, 136]. Additionally, metabolite concentrations in normal appearing brain tissue are better correlated with clinical scores than MRI lesion load [139].

As mentioned in the introduction section 1.7.4, MRSI has poor spatial resolution (typically around 1 cm^3) which limits the ability to differentiate among spatial features of interest, *e.g.,* between different tissue types. High spatial resolution MRSI cannot be acquired in the clinical routine with standard acquisition sequences as it requires a long scanning time.

Therefore, there is a need to design alternative, non-conventional approaches to create high-resolution MRSI-based metabolic images within a limited acquisition time. The acquisition time can be reduced by optimising the conventional phase encoded MRSI protocol [1, 118, 122] or by using parallel imaging reconstruction based techniques [47, 100, 115, 7] or combination of both [45], see [121] for a detailed review. Conventional MRSI protocols contain a series of periodic spatial and spectral encoding schemes. These could be optimised by reading out the spatial information during spectral acquisition. This optimised technique forms the basis of the echo planar spectroscopic imaging (EPSI) [122] which provides an improved spectral and spatial resolution compared to conventional MRSI protocols for fields up to 3T. Derived from the EPSI technique, spiral MRSI [1, 4, 19] uses a spiral trajectory in k-space, which allows faster acquisition than EPSI however at the cost of increased complexity in the data reconstruction. Parallel imaging reconstruction techniques are based on the principle of acquiring multiple undersampled k-space data using phased array coils. This undersampled data is then reconstructed into a single image using techniques like SENSE [127] or GRAPPA [71]. These techniques exploit the difference in spatial sensitivity among receiver coils to reconstruct an image from the undersampled data. Since optimised MRSI protocols like EPSI and parallel imaging are independent in nature, they can be combined to further reduce the acquisition time [45]. This reduction in the acquisition time generally comes at the cost of decreased SNR in the data. In spite of advances in speeding up the MRSI acquisition these techniques are yet not clinically feasible as the time required to attain the spatial resolution of conventional MRI sequences such as T1-weighted imaging still remains very high, although approaches that combine phase-encoding and EPSI [102] appear quite promising.

An ultra-high magnetic field strength such as 7T (used for non-routine research) offers more SNR, but the inherent limitations (*e.g.,* related to specific absorption rate (SAR) or B_0/B_1 homogeneity) have complicated high-resolution MRSI acquisition until the recent years. Free induction decay (FID) MRSI [18, 123, 161] with short acquisition delays was proposed to avoid SAR-intensive localisation schemes and SNR-loss due to $T2^*$ decay. The application of parallel imaging like SENSE [92, 186], GRAPPA [73] or CAIPIRINHA [162] allowed acceleration factors up to 9. Combining parallel imaging with a short TR of 200 ms, the acquisition of high-resolution MRSI within a 128×128 mm^2 field of view with a voxel volume of 23 μL and full slice coverage in around 10 minutes was successfully demonstrated in [74].

An alternative is to increase the MRSI resolution by super-resolution (SR) techniques. Super-resolution methods include k-space based reconstruction methods that improve the spectral quality of reconstructed high-resolution MRSI data using high-resolution spatial features, such as edges, from other imaging modalities, in particular T1-weighted MRI (for a detailed review see [89]). Super-resolution methods can be categorised into two sub-groups: methods based on a linear regression framework, or on a Bayesian framework. Linear regression based SR methods (LRSR) assume that the acquired data can be explained by a linear combination of a set of independent variables. Therefore, they aim at optimising the coefficients of these independent variables such that the error between the predicted and the actual measurements is minimised. One such method is spectral localisation by imaging (SLIM) [76] which assumes that the high-resolution MRSI consists of a linear combination of L anatomical compartments that are spectrally homogeneous. Although easy to implement, the assumption of spectral homogeneity may not hold true. To cope with this, two extensions have been proposed: generalised SLIM (GSLIM) [99] and SLIM with explicit B_0 field inhomogeneity compensation (BSLIM) [90]. GSLIM uses spatial Fourier harmonics to absorb any spatially dependent spectral variations and BSLIM assumes that the spectral variations are solely due to local (static) field inhomogeneity, therefore, requires an additional B_0 map for its correction. It has been observed that simple LRSR methods not always result in physically plausible solutions, therefore, an additional term commonly known as ‘regulariser’ is added to the optimisation problem so that the reconstruction remains well-behaved. For example, in [80], a local B-spline basis function is added as a regulariser to the GSLIM model to allow local intensity variations within each compartment. In [72], a smoothness regularisation term (controlled by pre-computed anatomically derived weight factors) has been added to penalise the local spectral variation between neighbouring voxels. The second sub-group of methods are based on Bayesian theory and model the reconstruction in the k-space as a likelihood function where the anatomical information acts as prior knowledge to estimate the optimised model parameters via the expectation-maximisation algorithm. For example in [8], the likelihood function consists of a combined spectral-spatial model where the tissue segmentations acts as prior information in estimating additional high frequencies in k-space. As an extension of this work, the likelihood model in [94] also addresses the spectral fitting problems and additionally uses prior information on the relationship between tissue segmentation and spatial metabolite distribution. These Bayesian based models are very complex in nature and have many variables to be optimised that often result in locally optimal solutions.

Contrary to the k-space based super-resolution approaches presented above, patch-based super-resolution (PBSR) methods have been used to reconstruct high-resolution T2 and diffusion weighted MR images [103, 104, 135, 34] in the spatial domain using a high-resolution T1-weighted image as reference. PBSR methods are based on the principle of image redundancy (ability to find similar patches in an image) and aim at finding similar voxels (defined using self similarity or using the high-resolution image (*e.g.,* T1-weighted image)) in the neighbourhood of the central voxel that could be used to guide the reconstruction of the central voxel. As opposed to classical interpolation methods (nearest neighbour, linear interpolation and B-splines), PBSR method provides better tissue contrast and thus yields better image quality reconstructions.

In view of good performance of PBSR methods in upsampling T2 and diffusion weighted MR images, we propose a multi parametric patch-based super-resolution method that acts directly on quantified MRSI metabolite maps from low-resolution MRSI data. In addition to the conventionally used high-resolution T1-weighted image, a FLAIR image is also used for lesion segmentation in multiple sclerosis subjects. The reconstruction process is guided by the intensities of T1-weighted and FLAIR images along with the brain segmentation, thus providing better estimation of brain tissue boundaries. To the best of our knowledge, a PBSR technique has never been applied before for upsampling MRSI-based metabolic images. We compare our method against the classical upsampling methods (nearest neighbour, linear interpolation and B-splines) on (a) an acetate image from a phantom, which contains different acetate proportions in different spatial locations, (b) simulated MS brain and quantified N-acetylaspartate images and (c) real images of MS patients with quantified N-acetylaspartate and myo-inositol images.

3.2 Methods

3.2.1 Method description

The patch-based super-resolution pipeline upsamples the low resolution quantified MRSI metabolite map using the high-resolution T1-weighted and FLAIR MR images. In the next sections, we describe the pipeline that has three steps:

1. a preprocessing step, including (a) MRI brain tissue segmentation, (b) metabolite quantification, (c) super-resolution initialisation.
2. a reconstruction step that estimates the metabolite concentration at each higher resolution MRSI voxel using the tissue segmentations along with image intensities of bias corrected T1-weighted and FLAIR images.
3. a mean correction step that rectifies the estimated metabolite concentration in every high-resolution voxel by taking into account a point spread function (PSF) and the error towards the corresponding lower level voxel's metabolite value.

The method iterates between step 2 and step 3 such that the corrected metabolite concentration map from previous iteration is used to initialise the metabolite concentration prior map for the current iteration. The convergence of our method is detected when the relative metabolite concentration difference between the current and previous iteration is negligible. It takes generally five iterations for the algorithm to converge. An overview of the pipeline is shown in figure 3.1.

Preprocessing

MRI brain tissue segmentation
The anatomical brain MR images are segmented into GM, WM, CSF and lesions using MS**metrix**-cross (see chapter 2). The method iteratively segments the T1-weighted image into GM, WM, and CSF, segments the WM lesions on the FLAIR image as an outlier to normal brain using Mahalanobis distance, and performs lesion filling in the T1-weighted image to improve tissue segmentation at next iteration. After convergence, soft segmentations of GM, WM and CSF are created together with binary lesion segmentation. In addition, bias corrected T1-weighted and FLAIR images are also produced.

Metabolite quantification
The acquired MRSI data, after Fourier transformation from k-space to the

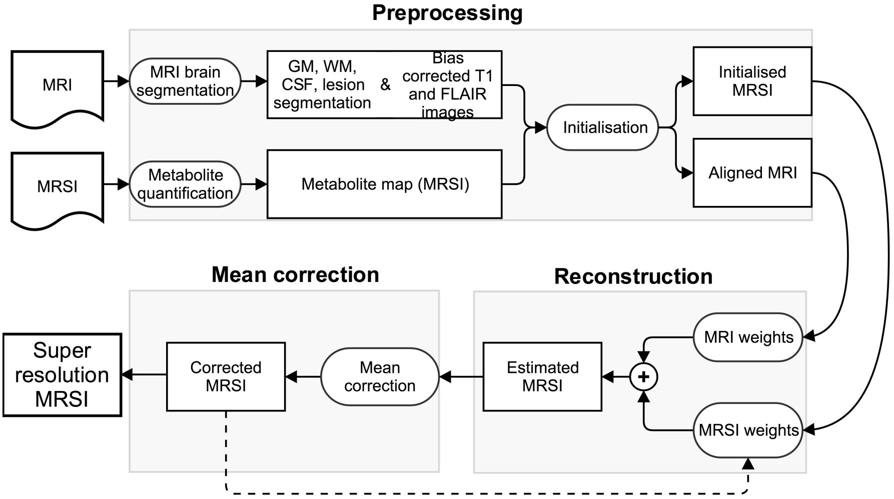


Figure 3.1: Schematic representation of patch-based super-resolution pipeline.

spatial domain, can be fitted in a voxel-by-voxel fashion using a simulated metabolite basis set, matching the acquisition parameters (*e.g.,* echo time) of the given signals. The fitting can be based either on time-domain MRSI signals, or after Fourier transformation to the frequency domain. In this work we used the SPID software[1] [173] and LCModel[2] [126] for metabolite quantification. The output consists of a set of metabolite images ('metabolite maps'), as many as metabolites in the basis set, and having the same spatial resolution as the input MRSI grid.

Super-resolution initialisation

The super-resolution pipeline is initialised by (1) upsampling the low resolution metabolite map using linear interpolation (scale factor of 2). This forms the starting value of the high-resolution MRSI metabolite map, (2) by aligning the MRI and MRSI images. As the MRI and MRSI images are acquired on the same scanner and in a sequential manner, their affine transformations are used to project the MRI image along with the tissue segmentations into the space of the upsampled MRSI image. The total affine transformation that relates these two image spaces is defined as:

$$Total_{affine} = MRSI^{-1}_{affine}.MRI_{affine} \tag{3.1}$$

---

[1] `SPID:http://homes.esat.kuleuven.be/~biomed/software.php`

[2] `LCModel:s-provencher.com/pages/lcmodel.shtml`

Patch-based super-resolution

Background

Let y be a noise-free low resolution metabolite map and x be the unknown corresponding high-resolution metabolite image defined over a high-resolution voxel space Ω. Theoretically,

$$y = H(x) \tag{3.2}$$

where H is a known blur and downsampling operator. In order to reconstruct x given y, the reconstruction model is typically formulated as an Euler Lagrange energy minimisation problem:

$$\hat{x} = arg\ min\{ \underbrace{||y - H(x)||^2}_{\text{least squares fitting}} + \lambda \underbrace{R(x)}_{\text{regulariser}} \} \tag{3.3}$$

where the parameter λ controls the amount of regularisation and the *R(x)* term preserves geometry and regularity and is defined as:

$$R(x) = \sum_{i \in \Omega} ||x_i - E(x_i|\zeta_i)||^2 \tag{3.4}$$

$E(x_i|\zeta_i)$ is the conditional reconstruction of x_i based on voxel intensity values in the 'search volume' ζ_i and is empirically defined as a cube of size $9 \times 9 \times 9$ voxels around the centre voxel x_i.

We assume that the term $E(x_i|\zeta_i)$ is consistent under strictly stationary conditions, *i.e.,* this stationary condition amounts to saying that, as the resolution of the image increases, there are many similar patches for all the details of the image. Also, when the regions move far from each other, their correlation decreases [23]. Both assumptions are approximately met in the case of MRSI. If $E(x_i|\zeta_i)$ is modelled under these assumptions, then $x_i \to E(x_i|\zeta_i)$ which means $R(x) \to 0$ [134]. Therefore, the two terms of the energy functional can be decoupled and asymptotically approximated as the reconstruction and mean correction steps of the iterative super-resolution algorithm.

```
Algorithm : SR($y$ = low res image; $x$ = starting value for high res image)

while $|x_i^t - x_i^{t-1}| > \epsilon, \ \forall\ x_i \in\ x$
    $Reconstruction : x_i^{t+1} = E(x_i|\zeta_i), \ \forall\ x_i \in\ x$
    $Mean\ correction : \hat{x_i}^{t+1} = x_i^{t+1} - \underbrace{(y_p - H(x_i^{t+1}))}_{\text{error}}$
```

where p is the index of the corresponding voxel at low resolution of which x_i is a part.

Reconstruction

The reconstruction term, $E(x_i|\zeta_i)$ in its general form is defined as:

$$E(x_i|\zeta_i) = \sum_{j \in \zeta_i} w(x_i, x_j) x_j \tag{3.5}$$

where the weight $w(x_i, x_j)$ defines the contribution of neighbourhood voxel x_j in the reconstruction of x_i. Mathematically, the general form of $w(x_i, x_j)$ is defined as [34]:

$$w(x_i, x_j) = \frac{1}{Z_i} e^{-\frac{1}{2N}\left(\frac{||\hat{N}_i - \hat{N}_j||}{h_i}\right)^2} \tag{3.6}$$

where Z_i is a normalisation constant such that $\sum_{j \in \zeta_i} w(x_i, x_j) = 1$, $\hat{N}_i$ is the intensity vector of the local neighbourhood N_i of length N, h_i is the standard deviation of neighbourhood voxel intensities in N_i. N_i is empirically defined as a cube of size $3 \times 3 \times 3$ voxels around the centre voxel x_i. From equation-3.6, we observe that if $\hat{N}_j$ is similar to $\hat{N}_i$, more weight is given to x_j in the reconstruction of x_i. Since we use both high-resolution MRI (T1-weighted and FLAIR) and low resolution MRSI, the weights are refined for both modalities taking into account the prior information:

Refining the weights for high-resolution MR images

$w_{MRI}(x_i, x_j)$ are defined such that the weights in (3.6) accommodate the prior knowledge on brain tissue segmentations that were down-sampled to match the MRSI resolution:

$$w_{MRI}(x_i, x_j) = \frac{1}{Z_i.K} \sum_{k}^{K} p_{i,k}.p_{j,k}.e^{-\frac{1}{2N}\left(\frac{||\hat{N}_i - \hat{N}_j||}{h_i}\right)^2} \tag{3.7}$$

where $k \in K = \{\text{GM, WM, CSF}\}$, $p_{.,k}$ denotes the probability that the voxel belongs to a particular tissue class k.

Refining the weights for low resolution MRSI

The $w_{MRSI}(x_i, x_j)$ weights have the form (3.6), and the only unknown term is h_i. Assuming white noise Gaussian distribution, h_i can be calculated using the pseudo residuals ϵ_i:

$$\epsilon_i = \sqrt{\frac{6}{7}}(x_i - \frac{1}{6} \sum_{n \in N_i^*} x_n) \tag{3.8}$$

where N_i^* is the 6-neighbourhood (4 in-plane and 2 out-of-plane neighbours) of x_i. Now,

$$h_i^2 = \frac{1}{M} \sum_{j \in N_i^*} \epsilon_j^2 \tag{3.9}$$

where M is the number of voxels in N_i^*.

After computing weights for both modalities, $E(x_i|\zeta_i)$ is defined similar to [135]:

$$E(x_i|\zeta_i) = \sum_{j \in \zeta_i} \Big\{ (1 - \alpha(x_i)) w_{MRI} + \alpha(x_i) w_{MRSI} \Big\} x_j \tag{3.10}$$

where $\alpha(x_i)$ is a weighing term between the MRI and MRSI weights. The choice of $\alpha(x_i)$ is application specific. For example, in case of non-pathological studies, the reconstruction process could be driven by MRI only, and therefore, $\alpha(x_i)$ can be set to zero. In case of pathological cases like MS, MRSI is more sensitive to pathological changes such as lesions compared to MRI. As we focus on MS subjects in this chapter, $\alpha(x_i)$ is defined using binary lesion segmentation which was obtained in the preprocessing step (see 3.2.1). In particular, the weight α is defined for each voxel x_i such that the $\alpha(x_i) = 1$ implies that it is a lesion and thus the reconstruction process is driven by MRSI only. Otherwise, it is driven by MRI (see equation 3.10).

Mean correction

In the mean correction step of the algorithm, the reconstructed values are first convolved with the PSF, which is typically a *sinc* function ($sinc(\pi x)$, where $x = \frac{1}{\text{nominal voxel size}} = \frac{\text{no of phase encoding steps}}{\text{field of view}}$) [121, 38]. Then, the average of the PSF corrected reconstruction values that compose the low resolution voxel y_p (of which x_i is part) must be close to the corresponding original value of the low resolution image. This corresponds to a sinc and boxcar operator for H in (3.2), but it could be easily replaced by a general smoothing and downsampling operator H.

3.2.2 Comparison with state-of-the-art methods

We compare our method's accuracy against three conventional interpolation techniques: nearest neighbour, linear interpolation and B-splines. The nearest neighbour (NN) function assigns the value of the new point with the closest old neighbour value. Linear interpolation (LIN) function linearly interpolates the new point between the old points. B-splines interpolation (BS) function uses piecewise cubic polynomials of degree 3 to interpolate the new point using four old points (two on each side of the new point).

3.2.3 Data

Phantom data

A cylindrical phantom (180 mm diameter, 168 mm height) is used, consisting of 7 small cylinders each having a diameter of 40 mm and depth of 150 mm, going through the thickness of the phantom as shown in figure 3.2. The first cylinder was kept empty for orientation purposes, followed by the second cylinder having acetate (Ace) concentration of 6 mM. Then, the Ace concentration was increased by 2 mM for each next cylinder resulting in a concentration of 16 mM for the last cylinder. The rest of the empty space in the phantom was filled with water. MR imaging was performed on a 3T whole body scanner (Philips Achieva, Best, the Netherlands). The protocol contained two sequences: 3D T1-weighted fast field echo sequence (TR = 6.6 ms, TE = 3.12 ms, FA = 9°, 180×180 mm^2 FOV, 3 axial slices, 0.8×0.8×10.0 mm^3 voxel resolution) and the ^{1}H MRSI point resolved spectroscopy (PRESS) sequence (TR = 1500 ms, TE = 120 ms, 300×300 mm^2 FOV, 3 axial slice, 10×10×10 mm^3 voxel resolution). MRI individual cylinder segmentation was performed manually using the Slicer[3] tool (version 4.3.1) and the super-resolution method was then performed at three different higher resolution voxel sizes: 5×5×5 mm^3, 2.5×2.5×2.5 mm^3 and 1.25×1.25×1.25 mm^3. As the phantom is homogeneous, we present in this chapter only the phantom data results at 1.25×1.25×1.25 mm^3 voxel resolution. MRSI data was quantified using the SPID software to obtain metabolite images for Ace.

Simulated brain datasets

The BrainWeb phantom datasets[4] [30] represent a human brain and contains three MS brain phantom images with mild, moderate and severe lesion volume. The protocol contained two sequences: 3D T1-weighted SFLASH sequence (TR = 18 ms, TE = 10 ms, FA = 30°, 181×181 mm^2 FOV, 217 sagittal slices, 1×1×1 mm^3 voxel resolution, 1% noise level and 20% RF field inhomogeneity) and custom simulated 3D FLAIR IR sequence (TR = 8000 ms, TE = 165 ms, TI = 1800 ms, FA = 90°, 181×181 mm^2 FOV, 217 sagittal slices, 1×1×1 mm^3 voxel resolution, 1% noise level and 20% RF field inhomogeneity). MRI tissue segmentation was performed using MS**metrix**-cross and these tissue segmentations were then used to simulate a high-resolution N-acetylaspartate (NAA) map (1×1×1 mm^3 voxel resolution) in the phantom brain images. In our study, the NAA concentration was chosen to be 30 (arbitrary units) in GM,

---

[3] `Slicer:www.slicer.org`

[4] `http://www.bic.mni.mcgill.ca/brainweb`

25 in WM and 20 in lesions. White Gaussian noise $\mathcal{N}(0, 2)$ is then added to the simulated high-resolution NAA map and the resulting image was downsampled using edge preserving Chebyshev type-I filter (n=31) to create a low resolution NAA map (2×2×2 mm^3 voxel resolution). MRI tissue segmentations were also downsampled by a scale factor of 2 to match the NAA map's low resolution.

Human datasets acquired in vivo

Five MS patients (4 females, 1 male, age 32–46) participated in the study at the High Field MR Centre, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria. This study was carried out in accordance with the recommendations of the 'International Conference on Harmonisation of Good Clinical Practice (ICH-GCP)', and the applicable Austrian legislation. The study was approved by the Institutional Review Board ethical committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki. All measurements were performed on a 7T whole body MR scanner (Magnetom, Siemens Healthcare, Erlangen, Germany) with a 32-channel receive coil array head coil (Nova Medical, Wilmington, MA, USA). The protocol contained four sequences: (1) 3D T1-weighted MPRAGE sequence (TR = 3800 ms, TE = 3.54 ms, FA = 9°, 230×230 mm^2 FOV, 208 axial slices, 0.7×0.7×0.7 mm^3 voxel resolution), (2) 3D SPACE FLAIR sequence (TR = 8000 ms, TE = 272 ms, FA = 160°, 215×215 mm^2 FOV, 160 axial slices, 0.8×0.8×0.8 mm^3 voxel resolution), (3) low resolution ^{1}H CAIPRINHA-accelerated phase-encoded FID-MRSI sequence [162] (TR = 600 ms, TE = 1.3 ms, 220×220 mm^2 FOV, single axial slice, 3.4×3.4×8.0 mm^3 voxel resolution, an acceleration factor of 6, 5 min measurement time), and (4) high-resolution ^{1}H CAIPRINHA-accelerated phase-encoded FID-MRSI sequence (TR = 200 ms, TE = 1.3 ms, 220×220 mm^2 FOV, single axial slice, 2.2×2.2×8.0 mm^3 voxel resolution, an acceleration factor of 4, 6 min measurement time). Both low and high-resolution MRSI data was acquired with WET (water suppression enhanced through T_1 effects) water suppression technique. MRI tissue segmentation was performed using MS**metrix**-cross and MRSI data was quantified using the LCModel (version 6.3-1) to obtain metabolite images for NAA and myo-inositol (myo-Ins).

3.2.4 Performance tests

In order to avoid the need of absolute quantification for the phantom data, we compare ratios between the estimated metabolite values obtained by each method in cylinders 5 and 2, and 7 and 3, respectively, as we know that the ground truth ratio is equal to 2.

For validating the accuracy on the simulated brain datasets and on the human datasets acquired in vivo, the global image similarity is measured with the structural similarity index (SSIM) [180] which is more compatible with the human visual assessment and is defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3.11}$$

where (μ_x, μ_y) and (σ_x,σ_y) are the respective means and standard deviations of image x and y. In our case, the metabolite concentrations are not zero or close to zero within brain, therefore, the stability constants c_1 and c_2 were set to zero.

Additionally, for both the simulated brain and the human brain acquired in vivo, statistical differences in the metabolite concentrations between the lesions and their neighbouring WM tissue are tested using Welsh's t-test [182] and the magnitude of this difference *i.e.,* the effect size is measured using Cohen's d [31].

3.3 Results

3.3.1 Accuracy on phantom dataset

Table 3.1 presents the quantitative results of Ace concentration ratio between cylinders 5 and 2, and 7 and 3 respectively, both at low resolution (acquired at 10×10×10 mm^3) and at high-resolution (reconstructed at 1.25×1.25×1.25 mm^3) for all methods. On average, PBSR and LIN seem to preserve better median metabolite ratio compared to other methods. However, the visual assessment for all methods (figure 3.2) shows that contrary to conventional interpolation methods, PBSR reduced the partial volume effects considerably by incorporating tissue segmentation information. Therefore, overall PBSR provides better results.

3.3.2 Accuracy on simulated brain datasets

Figures 3.3, 3.4 and 3.5 show the respective visual results for all methods on mild, moderate and severe MS subjects. On close inspection, it can be seen that PBSR maintains better tissue contrast compared to other methods. For PBSR, this contrast is also more evident between lesions and their surrounding white matter, both visually and quantitatively (see table 3.2). Moreover, the

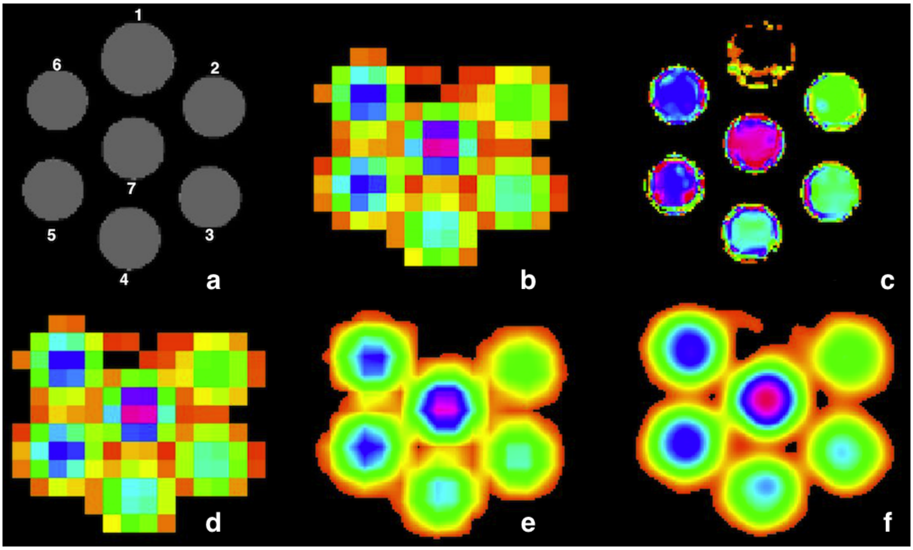


Figure 3.2: High-resolution Ace maps comparison for all methods. (a) Schematic representation of the phantom arrangement, (b) acquired low resolution Ace map. Reconstructed high-resolution Ace map for (c) patch-based super-resolution, (d) nearest neighbour, (e) linear interpolation, and (f) B-splines.

Table 3.1: Comparison between Ace concentration at low resolution (10×10×10 mm^3) and high-resolution (1.25×1.25×1.25 mm^3). Cylinders 5 and 7 have twice the Ace concentration of cylinders 2 and 3, respectively.

Ace ratio (median)	$\frac{cylinder\ 5}{cylinder\ 2}$	$\frac{cylinder\ 7}{cylinder\ 3}$
Low resolution	2.33	1.80
High-resolution		
PBSR	1.86	1.89
NN	2.25	1.6
LIN	1.75	2.0
BS	1.83	1.75

PBSR = patch-based super-resolution, NN = nearest neighbour, LIN = linear interpolation, BS = B-splines.

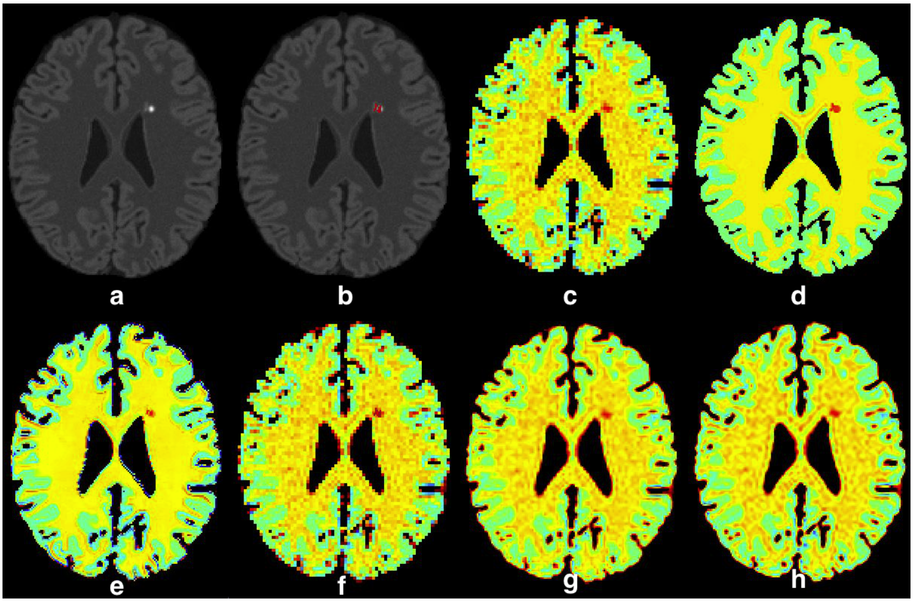


Figure 3.3: Results on simulated mild MS subject. (a) Bias corrected FLAIR image followed by (b) overlaid lesion segmentation, (c) simulated low resolution NAA map, (d-h) the high-resolution NAA maps: (d) simulated, (e) patch-based super-resolution, (f) nearest neighbour, (g) linear interpolation and (h) B-splines. The NAA concentration increases from red to blue.

median value in white matter surrounding lesions for PBSR is closest to the actual value considered for WM, 25, compared to other methods. For lesions, the median value for all methods is close to 20 with slightly better effect size for PBSR. SSIM score for all methods is very high with PBSR being marginally higher.

3.3.3 Accuracy on human datasets acquired in vivo

Figures 3.6 and 3.7 show the high-resolution NAA and myo-Ins maps respectively, for all methods on a representative example. In contrast to the results on the simulated brain datasets, visually it is very difficult to see a correlation between any method and the acquired high-resolution maps. However, quantitatively, a contrast between lesions and surrounding white matter can be computed (table 3.3). The acquired high-resolution shows a significant contrast between lesions and their surrounding WM, which is not seen at acquired low resolution. For NAA, this significant contrast can only be seen in PBSR. For myo-Ins, this

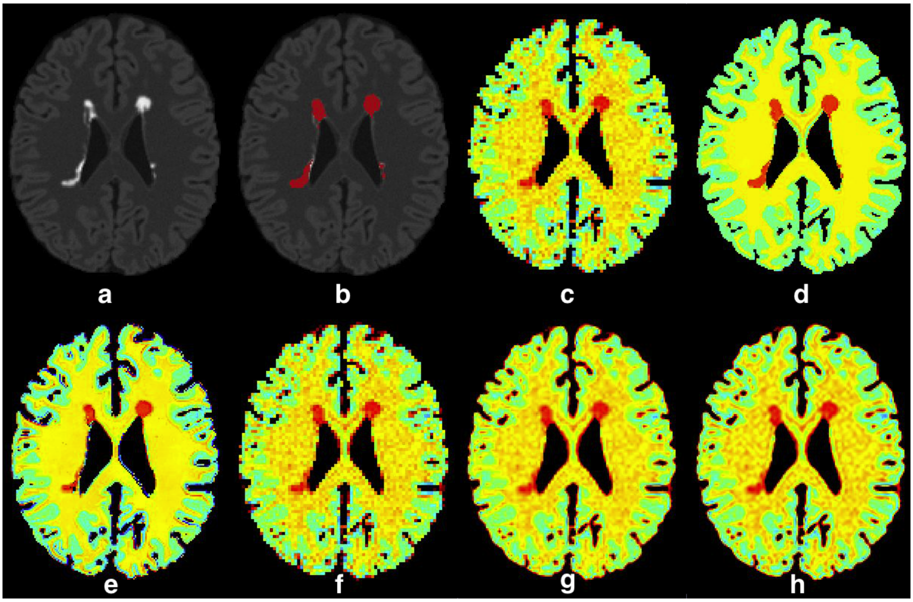


Figure 3.4: Results on simulated moderate MS subject. (a) Bias corrected FLAIR image followed by (b) overlaid lesion segmentation, (c) simulated low resolution NAA map, (d-h) the high-resolution NAA maps: (d) simulated, (e) patch-based super-resolution, (f) nearest neighbour, (g) linear interpolation and (h) B-splines. The NAA concentration increases from red to blue.

significant contrast is highest for LIN followed by BS, NN and PBSR. The average SSIM score is highest for PBSR for both NAA and myo-Ins compared to the other methods, which have equal average SSIM scores.

3.4 Discussion and conclusions

In this chapter, we presented a patch-based super-resolution method for upsampling the low resolution quantified metabolite maps. The method incorporates the high-resolution spatial information from T1-weighted and FLAIR MR images in guiding the reconstruction process. The method iteratively estimates and corrects for the metabolite concentration at a high-resolution. In contrast to the k-space based techniques where anatomically derived information is used to improve the spectral quality of high-resolution MRSI data, in our case, the method reconstructs each central voxel x_i using a weighted average of voxels that have similar tissue composition as the central voxel in the search

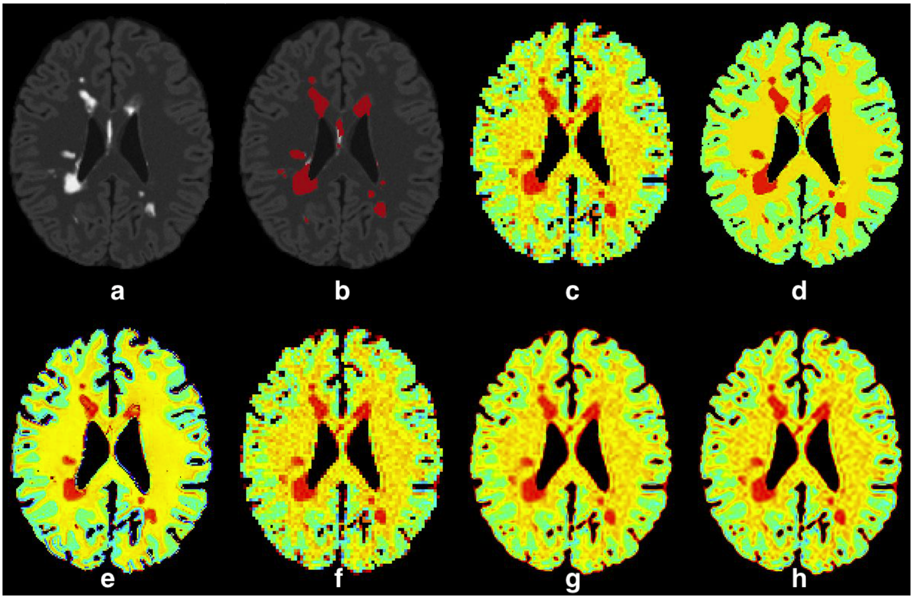


Figure 3.5: Results on simulated severe MS subject. (a) Bias corrected FLAIR image followed by (b) overlaid lesion segmentation, (c) simulated low resolution NAA map, (d-h) the high-resolution NAA maps: (d) simulated, (e) patch-based super-resolution, (f) nearest neighbour, (g) linear interpolation and (h) B-splines. The NAA concentration increases from red to blue.

volume. Moreover, as MRI and MRSI have complementary information, separate weights are defined for each modality which are regulated with a parameter $\alpha(x_i)$. Depending on the application, $\alpha(x_i)$ gives the flexibility to control the influence of each modality in the reconstruction process. In this chapter, when $\alpha(x_i) = 1$, the reconstruction is guided by MRSI only resulting in a smooth image and when $\alpha(x_i) = 0$, the reconstruction is guided by MRI alone resulting in a better tissue contrast compared to conventional interpolation techniques (see tables 3.2 and 3.3). The reconstruction quality is affected by the potential registration errors and the reliability of the input quantified metabolite map. Additionally, the method could not be expected to recover small scale features that were not sufficiently picked up by the low resolution measurement. On a similar note, the use of tissue segmentations to guide the reconstruction might affect metabolite values at tissue interfaces. In particular, through such an effect, some hyper intense voxels are introduced near the brain boundary, requiring further investigation. Finally, if the difference between the image resolution of MRI and MRSI images is big (scale factor more than 4), then the downsampled tissue segmentations for super-resolution process have high partial

Table 3.2: Quantitative measures for measuring the accuracy of all methods on simulated brain datasets.

			Mild		
	NWM$_{NAA}$	**Lesions$_{NAA}$**	**p-value**	**effect size**	**SSIM**
PBSR	25.26 (24.93, 25.62)	20.34 (19.33, 22.72)	3.07 e-08	2.9	0.93
NN	24.06 (23.7, 24.11)	20.07 (18.86, 21.02)	1.65 e-06	2.12	0.93
LIN	23.75 (22.83, 24.27)	20.46 (19.94, 21.7)	2.39 e-07	2.31	0.92
BS	23.7 (22.5, 24.45)	20.07 (19.2, 21.02)	7.20 e-07	2.18	0.92
			Moderate		
PBSR	25.26 (23.75, 25.5)	20.38 (19.56, 22.09)	1.61 e-18	1.11	0.93
NN	23.67 (21.65, 24.99)	20.1 (19.41, 21.65)	4.71 e-19	1.08	0.92
LIN	23.11 (21.04, 24.33)	20.37 (19.6, 21.66)	2.48 e-12	0.83	0.92
BS	23.37 (20.94, 24.5)	20.25 (19.53, 21.7)	8.43 e-14	0.89	0.92
			Severe		
PBSR	25.13 (22.12, 25.49)	20.47 (19.7, 21.66)	2.5 e-41	1.21	0.93
NN	22.98 (20.34, 24.83)	20.18 (19.75, 21.39)	9.54 e-36	1.11	0.92
LIN	22.94 (20.55, 24.35)	20.48 (19.77, 21.4)	1.2 e-32	1.04	0.92
BS	22.86 (20.56, 24.54)	20.28 (19.6, 21.11)	1.38 e-34	1.08	0.92

NWM$_{NAA}$ = NAA concentration in white matter surrounding lesions,
Lesions$_{NAA}$ = NAA concentration in lesions,
PBSR = patch-based super-resolution, NN = nearest neighbour, LIN = linear interpolation, BS = B-splines,
p-value = Welsh's t-test p-value for testing the statistical difference between NWM$_{NAA}$ and Lesions$_{NAA}$,
effect size = the magnitude of statistical difference, SSIM = structural similarity index.
NWM$_{NAA}$ and Lesions$_{NAA}$ are reported as median (first quartile - third quartile).

volume effect and thus do not carry much information on tissue type. That is why in this study, we limited the in-plane resolution of acquired low resolution MRSI to $\sim 3.4 \times 3.4$ mm^2. Our method can also be implemented in a multi-scale fashion. However, validating such high-resolution reconstructed image would be a problem in case of real datasets (*e.g.,* human datasets acquired in vivo) where acquiring such high-resolution image is not possible due to practical limitations. Indeed, multi-scale super-resolution method can easily be tested on a simulated data, and, as our results show better tissue contrast at the current high-resolution, we expect similar or better results at even higher resolution. Our main aim in this chapter was to validate the method on real datasets and thus we opted not to use the multi-scale super-resolution.

Among the methods proposed in the literature for super-resolution, our approach has some similarities to [135], which is also based on patch-based framework. In contrast with [135], which uses a neighbourhood averaging strategy to define MRI weights, we used brain tissue segmentations. In [135], $\alpha(x_i)$ defines the correlation between high-resolution and low resolution images and is adapted in each iteration. In our case, $\alpha(x_i)$ is defined using binarised lesion segmentation

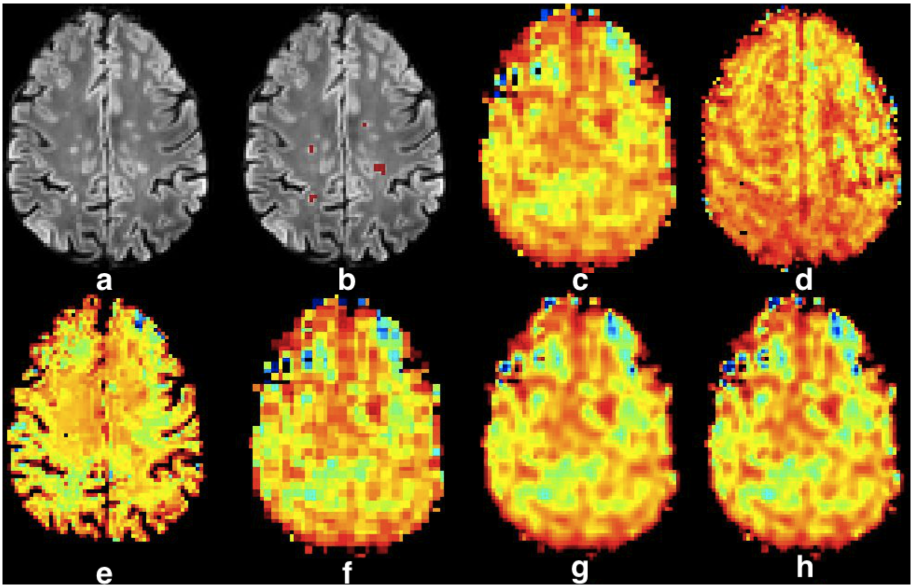


Figure 3.6: Qualitative accuracy performance of all methods on human dataset acquired in vivo for NAA metabolite. (a) Bias corrected FLAIR image followed by (b) overlaid lesion segmentation, (c) the acquired low resolution NAA map; (d-h) the high-resolution NAA maps: (d) acquired, (e) patch-based super-resolution, (f) nearest neighbour, (g) linear interpolation and (h) B-splines. The NAA concentration increases from red to blue.

and is fixed throughout the process.

One of the biggest problems that we encountered in this study was having ideal data for method validation. Challenges include (a) absence of high-resolution ground truth data, (b) lack of a publicly available dataset, and (c) difference between low and high-resolution acquired MRSI data. Acquiring high-resolution ground truth data in reality is not feasible because low metabolite concentration requires large voxel size to capture sufficient signal, and prolonged acquisition time. Although several group studies have reported metabolite concentration in normal and pathological cases like MS [139, 18], conclusions are not easily generalisable when it comes to individual patients. Lack of a standardised acquisition protocol for high spatial resolution MRSI hampers the creation of publicly available datasets against which state-of-the-art methods' performances can be compared. Moreover, most of the state-of-the-art super-resolution methods for MRSI improve MRSI data using sophisticated reconstruction techniques (k-space based), which make their reproducibility for comparison very difficult. However, from a methodological point of view, our method requires fewer variables to be optimised simultaneously which simplify the underlying

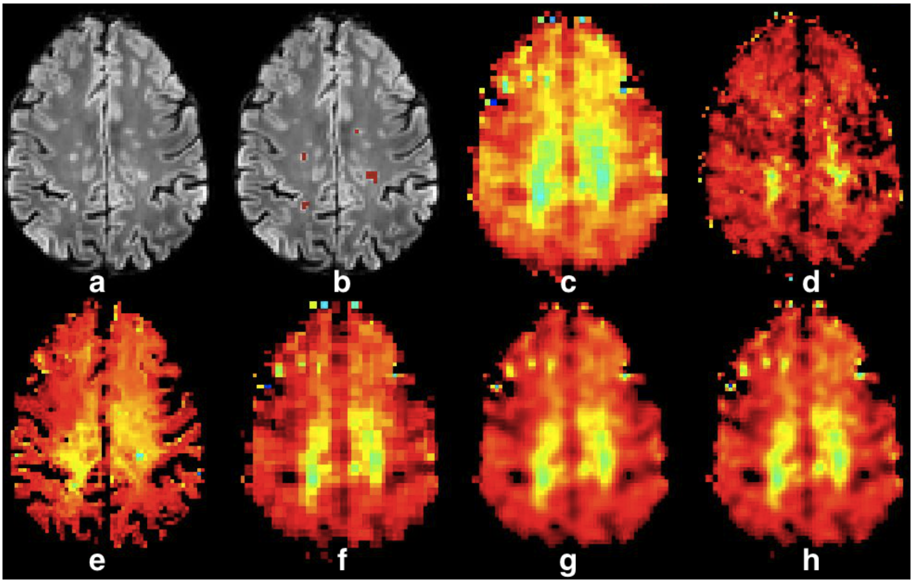


Figure 3.7: Qualitative accuracy performance of all methods on human dataset acquired in vivo for myo-Ins metabolite. (a) Bias corrected FLAIR image followed by (b) overlaid lesion segmentation, (c) the acquired low resolution myo-Ins map; (d-h) the high-resolution myo-Ins maps: (d) acquired, (e) patch-based super-resolution, (f) nearest neighbour, (g) linear interpolation and (h) B-splines. The myo-Ins concentration increases from red to blue.

optimisation problem. For example, in k-space based methods, the MRI spatial parameters (*e.g.,* bias field inhomogeneity, tissue class information) and the MRSI spectral parameters (*e.g.,* metabolite amplitudes, B_0 shifts, zero-order phase and lineshape parameters) have to be optimised simultaneously, which may result in a sub-optimal solution. In our case, we deal with the MRI spatial parameters and MRSI spectral parameters separately. Bias field correction and tissue class segmentation are performed by MSmetrix (3.2.1) and the interaction between neighbourhood voxels is exploited in the proposed super-resolution method. The spectral parameters are estimated by state-of-the-art metabolite quantification methods such as LCModel or SPID.

In human MRSI data acquired in vivo, the median metabolite concentrations acquired at low resolution are greater than those at higher resolution (see table 3.3), although the same quantification method (LCModel) and parameters have been used to quantify metabolic maps at both resolutions. This generates a degree of uncertainty in the interpretation of the performance analysis. This issue was not present in the simulated data, which explains a significant decrease in the SSIM score from simulated brain datasets' (see table 3.2) to human

Table 3.3: Accuracy results on human datasets acquired in vivo.

NAA					
	NWM$_{NAA}$	**Lesions**$_{NAA}$	**p-value**	**effect size**	**aSSIM**
ALR	15.18 (14.47, 16.87)	15.62 (13.39, 18.02)	0.8	-0.07	NA
AHR	13.41 (12.48, 15.0)	12.99 (10.99, 14.85)	0.03	0.31	NA
PBSR	16.85 (15.99, 17.6)	15.6 (13.82, 17.27)	2.0 e-05	0.64	0.18
NN	16.23 (14.82, 17.34)	15.75 (13.62, 17.61)	0.35	0.14	0.13
LIN	16.1 (14.93, 17.26)	15.71 (13.85, 17.35)	0.38	0.13	0.13
BS	16.15 (14.79, 17.33)	15.49 (13.64, 17.5)	0.3	0.15	0.13
myo-Ins					
ALR	10.8 (10.09, 12.69)	11.88 (10.9, 14.48)	5.0 e-02	-0.57	NA
AHR	6.03 (4.78, 7.86)	11.48 (9.13, 13.15)	4.3 e-16	-1.63	NA
PBSR	8.07 (6.78, 9.39)	10.59 (9.89, 11.79)	3.0 e-08	-1.04	0.2
NN	9.17 (7.77, 9.98)	11.97 (10.5, 13.35)	1.9 e-09	-1.15	0.18
LIN	9.11 (7.89, 10.11)	11.98 (10.82, 13.25)	4.7 e-11	-1.27	0.18
BS	9.12 (7.86, 10.15)	11.88 (10.72, 13.35)	1.2 e-10	-1.23	0.18

NWM$_{NAA}$ = NAA concentration in white matter surrounding lesions,
Lesions$_{NAA}$ = NAA concentration in lesions,
PBSR = patch-based super-resolution, NN = nearest neighbour, LIN = linear interpolation, BS = B-splines,
p-value = Welsh's t-test p-value for testing the statistical difference between NWM$_{NAA}$ and Lesions$_{NAA}$,
effect size = the magnitude of statistical difference, aSSIM = averaged structural similarity index over 5 subjects.
NWM$_{NAA}$ and Lesions$_{NAA}$ are reported as median (first quartile - third quartile).
ALR = acquired low resolution map, AHR = acquired high-resolution map.

datasets' acquired in vivo (see table 3.3). This decrease can also be explained by the fact that simulated low resolution MRSI data has twice the resolution of low resolution MRSI human datasets acquired in vivo, resulting in a loss of structural information. Finally, it may seem like the conventional interpolation methods provide better results than PBSR for myo-Ins (see table 3.3), however this is not completely true. Four out of five MS subjects in the human datasets acquired in vivo, have low lesion load in the acquired MRSI plane and the remaining subject has high lesion load. If this subject is removed, none of the methods show any statistical difference in the metabolite concentrations between the lesions and their neighbouring WM tissue for myo-Ins. However, this statistical difference still holds true for NAA. Tuning the parameter $\alpha(x_i)$ may address this issue.

In conclusion, we presented a patch-based super-resolution approach for upsampling the low resolution quantified metabolites maps using T1-weighted and FLAIR MR images. The proposed method preserves tissue contrast and structural information compared to conventional interpolation methods, and matches well with the trend of acquired high-resolution MRSI. These results suggest that the method has potential for clinically relevant neuroimaging applications.

Chapter 4

Two time points MS lesion segmentation

Abstract

Consistent MS lesion segmentation is important for monitoring of disease activity in a follow-up scan, where differences are clinically relevant. In this chapter, we present MS**metrix**-long: a joint EM framework for two time point white matter lesion segmentation. MS**metrix**-long takes as input a 3D T1-weighted and a 3D FLAIR MR image and segments lesions in three steps: (1) cross-sectional lesion segmentation of the two time points; (2) creation of difference image, which is used to model the lesion evolution; (3) a joint EM lesion segmentation framework that uses the output of step (1) and step (2) to provide the final lesion segmentation. MS**metrix**-long outperfoms MS**metrix**-cross in terms of accuracy and consistency in segmenting MS lesions. Also, MS**metrix**-long compares favourably with the publicly available longitudinal MS lesion segmentation algorithm of Lesion Segmentation Toolbox.

4.1 Introduction

Accurate and reliable lesion segmentation based on brain MRI scans is valuable for the diagnosis and monitoring of disease activity in patients with MS [17,

---

The work presented in this chapter has been published in: Jain, S., Ribbens, A., Sima, D.M., Cambron, M., De Keyser, J., Wang, C., Barnett, M.H., Van Huffel, S., Maes, F. and Smeets, D. Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework. *Frontiers in Neuroscience 10: 576* (2016), 11 pages.

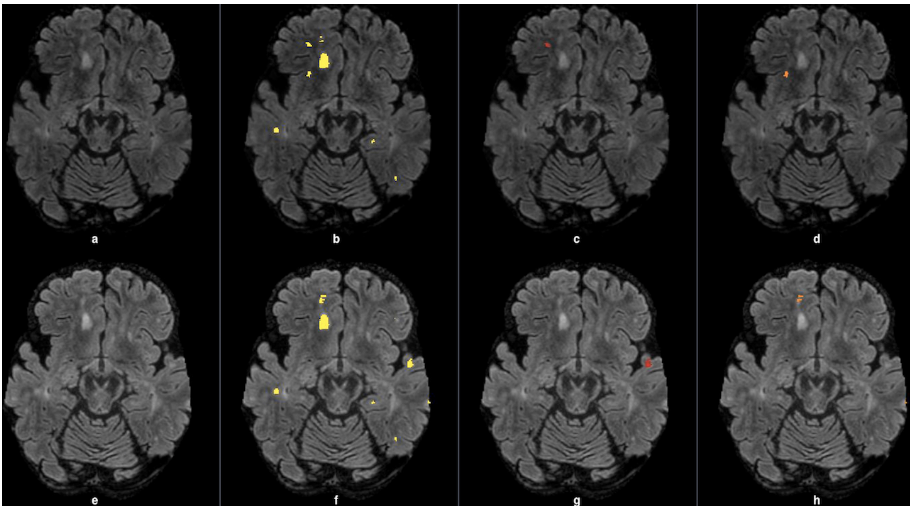


Figure 4.1: Bias corrected FLAIR images (a, e) followed by super-imposed lesion segmentations from: (b, f) the expert, (c) disappearing lesion, (d) shrinking lesion, (g) new lesion, and (h) enlarging lesion. The first row corresponds to time point 1 and the second row corresponds to time point 2.

41]. The availability of longitudinal MRI data permits an analysis of lesion evolution over time, a potential biomarker of disease progression and treatment efficacy. Figure 4.1 shows bias corrected FLAIR images of a MS subject scanned twice with an interval of approximately one year, along with the expert lesion segmentation followed by the lesion evolution *i.e.,* the new, disappearing, enlarging and shrinking lesions. Although expert manual delineation of lesions is considered as the gold standard, it is time consuming and often suffers from intra and inter observer variability [3]. To alleviate this problem, several automatic methods have been proposed in the literature to segment MS lesions. Interestingly, the vast majority of automatic methods are based on a single time point (cross-sectional) and relatively few methods take into account multiple time points (longitudinal) [61, 101]. Executing a cross-sectional method for each time point would indeed produce the longitudinal measures of interest, but such measures are less reliable as each time point is processed independently. Longitudinal methods incorporate both spatial and temporal information and are expected to be more reliable. Based on the underlying approach, longitudinal methods could be categorised in three different groups: change detection [20, 48, 124, 165, 130], 4D connectivity [14, 172] and outlier detection [2, 154] in multiple time points. Pre-processing of input MR images in these three groups is generally performed and consists of registration to a reference image or a common space, skull stripping, bias field correction and intensity normalisation.

Change detection methods primarily aim to detect MS activity by statistical analysis of image features or by measuring local volume variation. Statistical analysis can be performed in an unsupervised or supervised manner. Unsupervised approaches detect significant changes in the intensities between consecutive scans by either analysing the corresponding patches of two time points [20], or performing clustering on the extracted spatial and temporal features from longitudinal images [64, 124, 165]. The main drawback with unsupervised approaches is that they assume perfect registration and intensity normalisation. Supervised approaches learn the desired change from a training dataset; for instance, in [48], a random forest discriminative classifier was trained to learn relevant features (intensity, size and contextual information) related to new lesions and then use these features to segment them. The main drawback with this approach is that it often requires that the training dataset is large enough in order to capture all the distinctive features of the lesions to be segmented. To avoid the need for extracting image features, changes between consecutive images could be directly detected by measuring local volume variations. To this end, a Jacobian operator could be applied to the local deformation field obtained after non-rigid registration between the two time points. Although this approach has proven to be invariant to registration errors, it has given poor results for lesion segmentation [130].

Four-dimensional connectivity methods use voxel association in space and time to simultaneously segment and track lesion evolution. For example, [172] segments the lesions in two time points by clustering voxels that are both spatially and temporally adjacent to each other. The main disadvantage of this approach is that it often results in substantial false lesion segmentation. A more advanced method from the same family is based on spectral graph partitioning [14]. It constructs a 3D graph in which spatial pairwise affinities characterise lesions and background, and temporal affinities between adjacent time points represent lesion evolution direction. This graph is segmented into lesions and non-lesions via spectral clustering by maximising the force within-group attraction and between-group repulsion. The drawback of this approach is that it cannot discriminate between consistent artefacts and lesions.

Outlier detection methods are based on the fact that MS lesions are hyperintense on T2-weighted and FLAIR brain MRI scans and thus could be detected as an outlier to normal tissue class intensities distribution. For example, a joint EM based approach such as in [2] models the healthy brain tissue classes across the time points as a Gaussian mixture model (GMM) using a 4D (3D + time) intensity histogram. The parameters of the model are optimised via a modified version of the EM algorithm referred to as STREAM. After convergence, the lesions are extracted as outliers to healthy tissue classes using Mahalanobis distance and some prior information. In this approach the lesion segmentation is largely dependent on the choice of the Mahalanobis distance parameter and does not target lesion evolution, which is clinically relevant [2]. Another approach using outlier detection is based on the hidden Markov model (HMM) technique as in [154]. Initially, EM segments the first time point into different tissue classes including lesions, which are then manually corrected. Subsequently, using a lesion growth transition model and outlier detection sensor model, lesions are segmented in the following time points. The transition model enforces consistent lesion segmentation; however, it was validated only on simulations with exponential lesion growth.

In this chapter, we present MS**metrix**-long: an iterative white matter lesion segmentation method based on a joint EM framework that takes as input clinically acquired 3D T1-weighted and 3D FLAIR images of two time points. The proposed framework is fully automated, unsupervised and models the lesion evolution as GMM between two time points, thereby simultaneously segmenting new, enlarging, disappearing, shrinking and static lesions. The method is validated for accuracy and reproducibility on two different datasets that are representative for clinically feasible acquisition protocols.

4.2 Methods

4.2.1 Method description

The MS**metrix**-long pipeline analyses the MS lesions evolution between two time points based on 3D T1-weighted and 3D FLAIR image acquired at each time point. The pipeline has four steps:

1. cross-sectional analysis, that segments the individual time points into GM, WM, CSF and lesions.
2. FLAIR based difference image, which is created by subtracting the FLAIR images of both time points after bias correction, co-registration and intensity normalisation.
3. joint lesion segmentation, that aims to improve the individual time point lesion segmentation using the other time point information on tissue and lesion segmentation (initialised using step-1 results) and difference image obtained from step-2.
4. a pruning step, that refines the lesion segmentation obtained in the step-3 to eliminate non-lesion candidates.

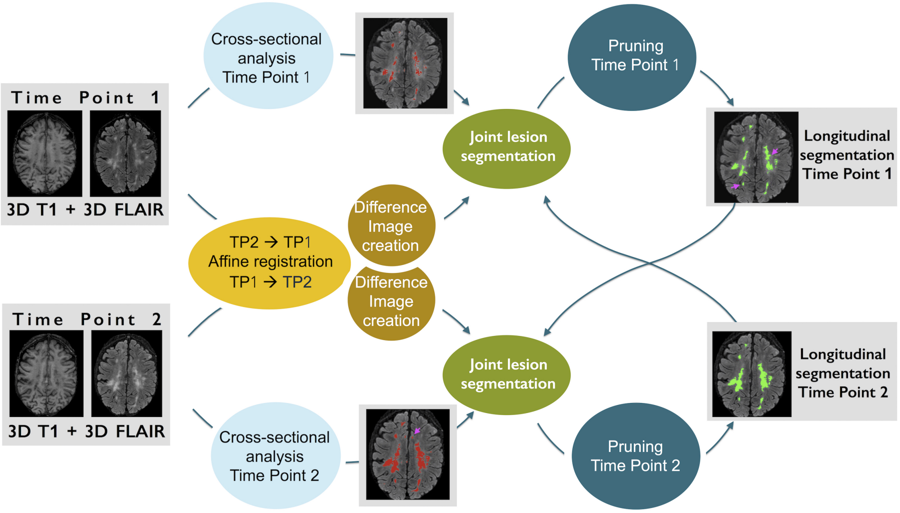


Figure 4.2: An illustrative example explaining the different steps of our method. The pink arrows in the longitudinal lesion segmentation at time point 1 show the recovered lesions using the second time point lesion segmentation and difference image information.

Figure 4.2 presents an illustrative explanation of these steps. Steps (3) and (4) are performed sequentially in both directions, by using one time point as reference and then the other. These steps are also iterated, by changing the input lesion segmentation used as prior. Only for the first iteration, the lesion segmentations priors come from the cross-sectional pipeline in step-3, while from the second iteration onwards lesion segmentations from previous iteration are used to initialise the lesion priors for the current iteration. The convergence of our method is decided when the relative lesion segmentation difference between the current and previous iteration is negligible. It takes generally three iterations for the algorithm to converge. The following sections explain the different steps in more detail.

Cross-sectional analysis

Image segmentation is performed independently for each time point using the cross-sectional pipeline referred to as MS**metrix**-cross (see chapter 2). The cross-sectional method iteratively segments the T1-weighted image into GM, WM and CSF, segments the WM lesions on the FLAIR image as an outlier to normal brain using Mahalanobis distance, and performs lesion filling in the T1-weighted image to improve tissue segmentation at next iteration. After convergence, segmentations of WM, GM, CSF and lesions are created. In addition, bias corrected T1-weighted and FLAIR images are also produced. The segmentation tasks of the MS**metrix**-cross are optimised using an EM algorithm [177] as implemented in NiftySeg [26].

FLAIR based difference image

A FLAIR based difference image is created by image co-registration and intensity normalisation. Image co-registration is performed using affine registration, which comprises a rigid registration based on the whole T1-weighted image, followed by a skull based affine registration to avoid small scaling differences, and a final whole brain rigid registration [151]. The rigid registration and skull based affine registration use an inverse consistent registration algorithm [107]. Subsequently, the GM, WM, CSF, lesion segmentation and the bias corrected FLAIR images obtained from the cross-sectional analysis are propagated using the final affine transformation. The matched bias corrected FLAIR images are then corrected for differential bias field as described in [98]. Subsequently, the differential bias field corrected images are intensity normalised using a cumulative histogram matching technique [27] with the image of time point 1 as reference. A FLAIR based difference image is now created in time point 1 space. To avoid bias towards a specific time point, a second difference image is created, using time point 2 space as reference.

Joint lesion segmentation

We now define a model that aims at joint tissue class label segmentation of the images of both time points (see the blocks denoted by 'Joint lesion segmentation' in figure 4.2). The observed variables for the model are FLAIR image intensities of both time points as well as the corresponding FLAIR based difference image. The hidden variables for the model are tissue class labels for which prior knowledge is initially provided in the form of tissue class segmentations from MS**metrix**-cross. Optimisation of the model results in a lesion segmentation at each time point that is based on information from both images. As mentioned above, to avoid a bias towards a specific time point, the joint lesion segmentation is performed separately in both directions. We now describe the notations, variables and assumptions used, followed by the model definition.

Notations, variables and model assumptions

We assume that image 1, image 2 and difference image are co-registered and have the same number of voxels, N_J. Let $I_1 = \{y_j^{(1)} \mid \text{j} \in \{1, .., N_J\}\}$, $I_2 = \{y_j^{(2)}$

| j $\in \{1, .., N_J\}\}$ and $D = \{y_j^{(D)}$ | j $\in \{1, .., N_J\}\}$ be the set of image intensities with $y_j^{(1)}$, $y_j^{(2)}$ and $y_j^{(D)}$ being the intensities corresponding to voxel j in image 1, image 2, and difference image respectively.

We assume identical tissue classes in both images and their total number denoted by N_K. Let $k = \{1, .., N_K\}$ be the set of tissue classes with $k^{(1)}$ and $k^{(2)}$ denoting a tissue class index for image 1 and image 2, respectively such that $k^{(1)}$ and $k^{(2)} \in \{1, .., N_K\}$. The tissue class label in image 1 is defined as: $L_1 = \{l^{(1)}_{j,k^{(1)}}$ | j$\in\{1, .., N_J\}$, $k^{(1)} \in \{1, .., N_K\}\}$; $l^{(1)}_{j,k^{(1)}}$ is 1 if voxel j has tissue class label $k^{(1)}$ in I_1, 0 otherwise; $l^{(1)}_{j,k^{(1)}}$ is subjected to the constraint: $\sum\limits_{k^{(1)}} l^{(1)}_{j,k^{(1)}} = 1$. Similarly, the set of tissue class labels in image 2 are denoted by L_2 and defined in the same way as for L_1.

We now specify our model assumptions. A Gaussian mixture model is used on the image intensities of each time point where a Gaussian model is used for each tissue class k. Let $\{\mu^{(1)}_{k^{(1)}}, \sigma^{2(1)}_{k^{(1)}}\}$ be the mean and variance of Gaussian model parameters for tissue class $k^{(1)}$ and let θ_1 denote the Gaussian mixture model parameters for the intensities of image 1. The probabilistic model for image 1 can be defined as:

$$P(I_1|L_1, \theta_1) = \prod_{j,k^{(1)}} \mathcal{N}\left(y_j^{(1)}|\mu^{(1)}_{k^{(1)}}, \sigma^{2(1)}_{k^{(1)}}\right)^{l^{(1)}_{j,k^{(1)}}} \tag{4.1}$$

Analogously, the probabilistic model for image 2 can be defined as:

$$P(I_2|L_2, \theta_2) = \prod_{j,k^{(2)}} \mathcal{N}\left(y_j^{(2)}|\mu^{(2)}_{k^{(2)}}, \sigma^{2(2)}_{k^{(2)}}\right)^{l^{(2)}_{j,k^{(2)}}} \tag{4.2}$$

We make the underlying assumption that the ‘difference image’ might be independently generated as an image that captures anatomical changes including new lesions or atrophy. The image created by subtracting image 1 from image 2 or vice-versa (after intensity normalisation) is one such instance of the difference image. The intensity model of image 1 and image 2 can therefore be improved by including a transition model defined on the difference image. As our method focuses on two time points WM lesion segmentation, we only model the transformations between WM and lesions. We assume that the difference image has three different transformations: ‘static’, ‘growth’ and ‘shrinkage’. The static transformation class is defined as a set of voxels in the difference image that are either labeled as WM in both images or lesions. The growth transformation class (describing the new and enlarging lesions) is defined as a set

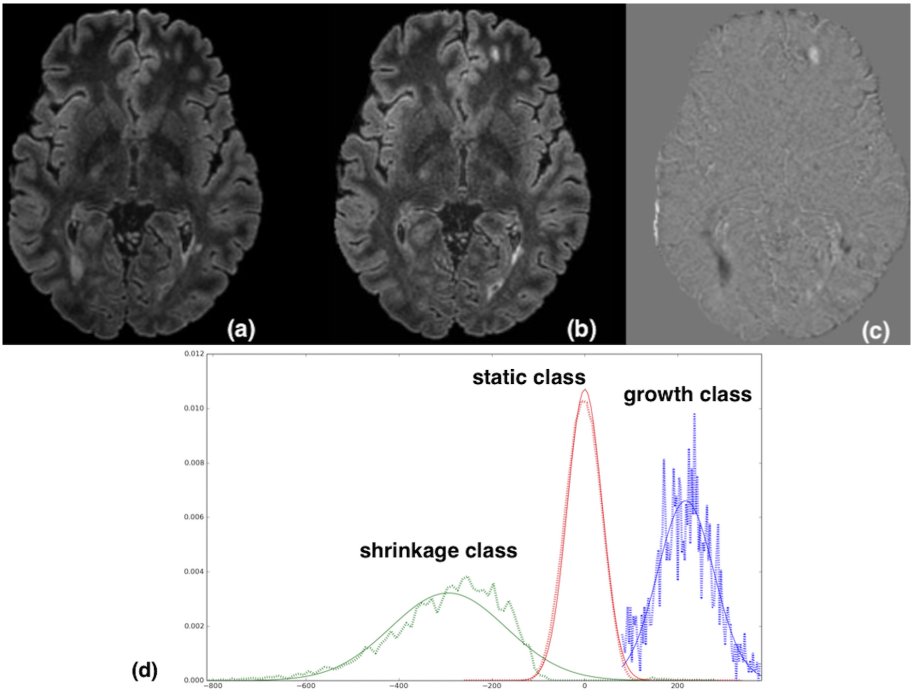


Figure 4.3: (a) Normalised FLAIR image of time point 1, (b) FLAIR image of time point 2, (c) difference image (b-a), (d) histograms of difference image classes with corresponding Gaussian fitting, normalised per class. Note the artifactual difference values at the brain contour (due to subtle differences in brain mask extraction) are excluded by only including WM voxels in the analysis.

of voxels in the difference image that are labelled as WM in image 1 and lesion in image 2. The shrinkage transformation class (describing the disappearing and shrinking lesions) is defined as a set of voxels in the difference image that are labelled as lesion in image 1 and WM in image 2. For all other possible tissue transformations from image 1 and image 2 a uniform distribution is assumed. Figure 4.3 shows an illustrative example of the difference image and the histograms of its classes with corresponding Gaussian fitting.

Under these assumptions, a Gaussian mixture model for the difference image intensities is used where each transformation class is modelled as Gaussian:

$$P(D|L_1, L_2, \zeta) = \begin{cases} \prod_{j,k^{(1)},k^{(2)}} \mathcal{N}\left(y_j^{(D)} | \mu_{k^{(1)},k^{(2)}}^{(D)}, \sigma_{k^{(1)},k^{(2)}}^{2(D)}\right)^{l_{j,k^{(1)}}^{(1)},\ l_{j,k^{(2)}}^{(2)}} & \text{, with } k^{(1)} \text{ and } k^{(2)} \in \{\text{WM, lesion}\} \\ \propto\ 1 & \text{, } \forall \text{ other } k^{(1)} \text{ and } k^{(2)} \text{ combinations.} \end{cases} \tag{4.3}$$

where the parameter $\zeta = \{\theta_{static},\, \theta_{growth},\, \theta_{shrinkage}\}$ with

$$\theta_{static} = \{\mu^{(D)}_{static}, \sigma^{2(D)}_{static}\}, \qquad \text{if} \begin{cases} k^{(1)} &= k^{(2)} = \text{WM} \;\; \text{or} \\ k^{(1)} &= k^{(2)} = \text{lesion} \end{cases} \tag{4.4}$$

$$\theta_{growth} = \{\mu^{(D)}_{growth}, \sigma^{2(D)}_{growth}\}, \qquad \text{if} \begin{cases} k^{(1)} &= \text{WM} \;\; \text{and} \\ k^{(2)} &= \text{lesion} \end{cases} \tag{4.5}$$

$$\theta_{shrinkage} = \{\mu^{(D)}_{shrinkage}, \sigma^{2(D)}_{shrinkage}\}, \qquad \text{if} \begin{cases} k^{(1)} &= \text{lesion} \;\; \text{and} \\ k^{(2)} &= \text{WM} \end{cases} \tag{4.6}$$

Finally, we assume that we have no prior knowledge on the relationship of the tissue class labels between both images which is generally provided by a probabilistic atlas. Therefore, we define the prior probabilities independently for each image using the tissue class segmentations from our cross-sectional model. The prior probabilities on tissue class labels for the image 1 and image 2 are denoted by $P(L_1)$ and $P(L_2)$, and are defined as:

$$P(L_1) = \prod_{j,k^{(1)}} \left(\alpha^{(1)}_{j,k^{(1)}}\right)^{l^{(1)}_{j,k^{(1)}}} \quad , \quad \mathrm{P}(L_2) = \prod_{j,k^{(2)}} \left(\alpha^{(2)}_{j,k^{(2)}}\right)^{l^{(2)}_{j,k^{(2)}}} \tag{4.7}$$

where $\alpha^{(1)}_{j,k^{(1)}}$ and $\alpha^{(2)}_{j,k^{(2)}}$ are the tissue class segmentations of image 1 and image 2 respectively. For the first iteration of joint lesion segmentation, these tissue class segmentations comes from the cross-sectional method. From the second iteration, the obtained lesion segmentations from previous iteration are used to initialise the lesion priors for current iteration.

The model

Under these assumptions, the joint probabilistic model could be formulated as follows:

$$P(I_1, I_2, D, L_1, L_2, \gamma) = P(I_1|L_1, \theta_1).\ P(I_2|L_2, \theta_2).\ P(D|L_1, L_2, \zeta).\ P(L_1).\ P(L_2) \tag{4.8}$$

where $\gamma = \{\theta_1, \theta_2, \zeta\}$. Our model is optimised by the maximum a posteriori (MAP) problem shown in equation 4.9. Since the knowledge of tissue class labels (hidden variables) helps in finding the model parameters and vice-versa, we reformulate our MAP problem as presented in equation 4.10.

$$\hat{\gamma}_{\mathrm{MAP}} = \underset{\gamma}{argmax} \ln \ P(\gamma|I_1, I_2, D) = \underset{\gamma}{argmax} \ln \ P(I_1, I_2, D, \gamma) \tag{4.9}$$

$$= \underset{\gamma}{argmax} \ln \sum_{L_1, L_2} P(I_1, I_2, D, L_1, L_2, \gamma) \tag{4.10}$$

$$\geq \underset{\gamma}{argmax} \sum_{L_1, L_2} P(L_1, L_2|I_1, I_2, D, \overline{\gamma}) \ln \ \frac{P(I_1, I_2, D, L_1, L_2, \gamma)}{P(L_1, L_2|I_1, I_2, D, \overline{\gamma})} \tag{4.11}$$

Finally, a lower bound of our model is derived using Jensen's inequality and optimised by the EM algorithm. The Q-function or the E-step can now be written as:

$$Q(\gamma|\overline{\gamma}) = E_{L_1, L_2|I_1, I_2, D, \overline{\gamma}} \left[\ln \ P(I_1, I_2, D, L_1, L_2, \gamma) \right] \tag{4.12}$$

with the joint posterior distribution $P(L_1, L_2|I_1, I_2, D, \overline{\gamma})$, which can be formulated in every voxel j for tissue class $k^{(1)}$ and $k^{(2)}$ as:

$$p_{j,k^{(1)},k^{(2)}} = \frac{p(y_j^{(1)}, y_j^{(2)}, y_j^{(D)}, l_{j,k^{(1)}}^{(1)} = 1, l_{j,k^{(2)}}^{(2)} = 1, \overline{\gamma})}{\sum\limits_{k^{(1)},k^{(2)}} p(y_j^{(1)}, y_j^{(2)}, y_j^{(D)}, l_{j,k^{(1)}}^{(1)}, l_{j,k^{(2)}}^{(2)}, \overline{\gamma})} \tag{4.13}$$

The sum over all possible tissue classes $k^{(2)}$ of the joint posterior gives us the soft tissue class segmentations at time point 1, referred to as $p_{j,k^{(1)}}$. Similarly, the sum over all possible tissue classes $k^{(1)}$ of the joint posterior gives us the soft tissue class segmentations at time point 2, referred to as $p_{j,k^{(2)}}$.

$$p_{j,k^{(1)}} = \sum_{k^{(2)}} p_{j,k^{(1)},k^{(2)}} \tag{4.14}$$

$$p_{j,k^{(2)}} = \sum_{k^{(1)}} p_{j,k^{(1)},k^{(2)}} \tag{4.15}$$

In the M-step, a new set of values for model parameter γ is computed by maximising the Q-function. For the Gaussian parameters of the tissue class $k^{(1)}$ we find:

$$\mu_{k^{(1)}}^{(1)} = \frac{\sum\limits_j p_{j,k^{(1)}} y_j^{(1)}}{\sum\limits_j p_{j,k^{(1)}}} \tag{4.16}$$

$$\sigma^{2(1)}_{k^{(1)}} = \frac{\sum\limits_j p_{j,k^{(1)}} \left[y^{(1)}_j - \mu^{(1)}_{k^{(1)}}\right]^2}{\sum\limits_j p_{j,k^{(1)}}} \tag{4.17}$$

In a similar way, the Gaussian parameters of the tissue class $k^{(2)}$ $\{\mu^{(2)}_{k^{(2)}}, \sigma^{2(2)}_{k^{(2)}}\}$ could be derived.

For the difference image parameter $\zeta = \{\mu^{(D)}_{k^{(1)},k^{(2)}}, \sigma^{2(D)}_{k^{(1)},k^{(2)}}\} = \{(\mu^{(D)}_{static}, \sigma^{2(D)}_{static}), (\mu^{(D)}_{growth}, \sigma^{2(D)}_{growth}), (\mu^{(D)}_{shrinkage}, \sigma^{2(D)}_{shrinkage})\}$, the general form of mean and variance for every transformation class can be written as:

$$\mu^{(D)}_{k^{(1)},k^{(2)}} = \frac{p_{j,k^{(1)},k^{(2)}}\ y^{(D)}_j}{\sum\limits_j p_{j,k^{(1)},k^{(2)}}} \tag{4.18}$$

$$\sigma^{2(D)}_{k^{(1)},k^{(2)}} = \frac{p_{j,k^{(1)},k^{(2)}} \left[y^{(D)}_j - \mu^{(D)}_{k^{(1)},k^{(2)}}\right]^2}{\sum\limits_j p_{j,k^{(1)},k^{(2)}}} \tag{4.19}$$

where $k^{(1)}$ and $k^{(2)}$ $\in$ {WM, lesion} and depending on the $k^{(1)}$ and $k^{(2)}$ combination as mentioned in equations-[4.4, 4.5, 4.6] mean and variance of each transformation class can be derived.

Pruning

The soft lesion segmentations obtained from the E-step of the joint EM algorithm are pruned to eliminate non-lesions (such as partial volume effects, artefacts) that share intensities and locations with the potential lesions. Thereto, a priori information on the appearance, location and volume of lesions is incorporated: (1) the lesion intensities should be hyper-intense compared to the WM intensities on bias field corrected FLAIR image, (2) the lesions are in the WM region, and (3) the lesion needs to have a minimum number of 5 adjacent voxels (empirically determined) to avoid spurious lesion detection. The hyper-intensity is defined as the mean plus two times the standard deviation of WM intensities. The intensities and location of WM region are computed using the WM segmentation from the MS**metrix**-cross pipeline. In addition, a priori defined binary mask (defined in the MNI space and consisting of the cerebral cortex and WM in-between the ventricles) is warped to the subject space to remove lesion candidates from these regions that are likely to result in a false lesion segmentation. After the pruning, the soft lesion segmentations are binarised using a threshold of 0.9 (empirically determined) on the posterior probabilities.

4.2.2 Comparison with state-of-the-art methods

We compare MS**metrix**-long pipeline[1] with the MS**metrix**-cross pipeline to know the gain over the cross-sectional method. Furthermore, we also compare against the longitudinal pipeline of the LST software package [142], version 2.0.12, which is implemented in SPM12[2]. The longitudinal pipeline of LST, which is referred to as LST-long in this chapter, performs individual time point lesion segmentation using the lesion growth algorithm described in [142]. The obtained lesion segmentation maps of different time points are coregistered to the baseline scan and are corrected by comparing the relative differences of FLAIR intensities in all lesion maps to produce the final lesion segmentation at each time point (see LST documentation[3]).

For comparison, all three methods were executed on the same datasets and default parameter settings were used. Thus, no parameter tuning was performed at dataset or subject level.

4.2.3 Data

Dataset 1 contains scans from 12 relapsing remitting MS patients on a GE 3T scanner (Discovery MR750), each scanned twice at an interval of approximately one year. Therefore, the sample size of dataset 1 equals 24. Each time point contained two a 3D sequences: a CUBE FLAIR (TR: 8000 ms, TE: 165 ms, TI: 2179 ms) and a 3D T1-weighted IR-FSPGR sequence (TR 7.2 ms, TE 450 ms, TI 2.8 ms). Both 3D sequences have voxel resolution close to 1 mm^3. Expert WM lesion segmentations were created on the baseline FLAIR scan by the experienced neuro-imaging analyst using JIM software tool[4], version 6.0. For follow-up scans, baseline lesion segmentation was overlaid on rigidly registered follow-up scan at the beginning, and then the lesion segmentation was adapted according to lesion activities. This study was reviewed and conducted within the guidelines set out in the National Statement on Ethical Conduct in Human Research (2007) in Australia, and approved by University of Sydney Human Research Ethics Committee. All subjects gave written informed consent.

The second dataset, dataset 2 contains scans from 10 MS patients scanned twice, with re-positioning (time interval between two scans is 5 $\sim$10 minutes), on each of three different 3T scanners from GE (Discovery MR750w), SIEMENS

---

[1] MS**metrix**-long pipeline presented here is the version 1.4 of the clinical product – MS**metrix**.

[2] `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`

[3] `LST,www.statistical-modelling.de/lst.html`

[4] `JIM,http://www.xinapse.com`

(Skyra) and PHILIPS (Achieva). This dataset is the same as the dataset 2 used in chapter 2, section 2.2.3, table 2.1.

4.2.4 Performance tests

Accuracy and reproducibility assessment

The agreement between the expert segmentation and automatic methods on dataset 1 is evaluated at three levels: voxel-by-voxel, lesion-wise and volumetric. Voxel-by-voxel metric includes the Dice similarity index. The lesion-wise metrics include lesion-wise true positive rate, false positive rate, F1 score, absolute lesion change difference and Pearson correlation coefficient (r value). Lesion-wise true positive rate (LTPR) is defined as the ratio of the total number of lesions where the expert reference and the automatic segmentation intersect to the total number of lesions in the expert reference segmentation. Lesion-wise false positive rate (LFPR) is defined as the ratio of the total number of lesions that are present only in the automatic segmentation to the total number of lesions in the automatic segmentation. Lesion-wise F1 score is defined as the harmonic mean of LTPR and LFPR. Absolute lesion-wise change difference is defined as the absolute difference between the overall lesion-wise change (number of new lesions minus number of disappearing lesions) in the expert lesion segmentation and the automatic segmentation. In this chapter, we consider new, disappearing, enlarging and shirking lesions that have size more than 20 voxels and at least one slice which encompasses the lesion presents a minimum of 5 lesion voxels.

Volumetric metrics measure the total lesion volume agreement and consist of the r value and the absolute volume difference. The absolute volume difference is computed as the absolute difference between the total volume reported by the expert reference segmentation and the corresponding value derived from the automatic method.

The reproducibility of the method is evaluated on dataset 2 by the Dice similarity index of the lesion segmentations at both times points. Moreover, the estimated number of new lesions and the absolute total lesion volume difference is also calculated between time points, which are both expected to be zero in this test-retest scenario.

To determine if there is a statistical difference between MS**metrix**-long and LST-long and between MS**metrix**-cross and MS**metrix**-long methods' performance, two tailed paired Wilcoxon signed-rank test is performed.

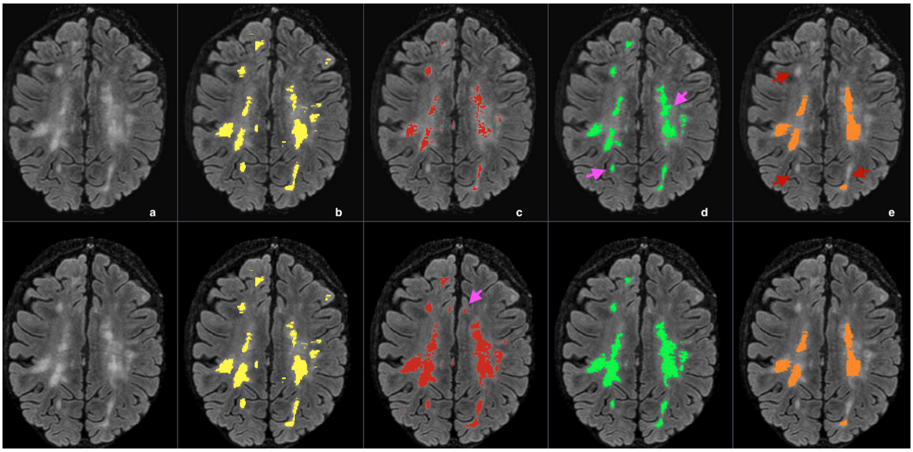


Figure 4.4: Bias corrected FLAIR image (a) followed by super-imposed lesion segmentation from: (b) expert segmentation, (c) MS**metrix**-cross, (d) MS**metrix**-long, and (e) LST-long. The first row corresponds to the lesion segmentation of time point 1 and the second row corresponds to the lesion segmentation of time point 2. Pink arrows specify places where MS**metrix**-long has improved in accuracy over MS**metrix**-cross and red arrows indicate regions where LST-long has missed lesions.

4.3 Results

4.3.1 Accuracy results on dataset 1

Figure 4.4 shows a representative example of lesion segmentation obtained by MS**metrix**-cross, MS**metrix**-long and LST-long on a patient from dataset 1. By comparing against expert delineations, it can be observed that MS**metrix**-long has improved in accuracy over MS**metrix**-cross and that LST-long has missed lesions.

The volumetric correlation of MS**metrix**-long and LST-long to the expert reference segmentation can be visualised in figure 4.5. MS**metrix**-long has a better correlation ($r = 0.96$) with expert reference segmentation compared to LST-long ($r = 0.88$).

Table 4.1 summarises the cross-sectional lesion segmentation performance of MS**metrix**-cross, MS**metrix**-long and LST-long on dataset 1 ($n = 24$) in a quantitative way. MS**metrix**-long has improved over MS**metrix**-cross in the median Dice, F1 score and LFPR. Compared to LST-long, MS**metrix**-long has a higher median Dice, F1 score, LTPR and r value, together with lower LFPR and absolute lesion volume difference.

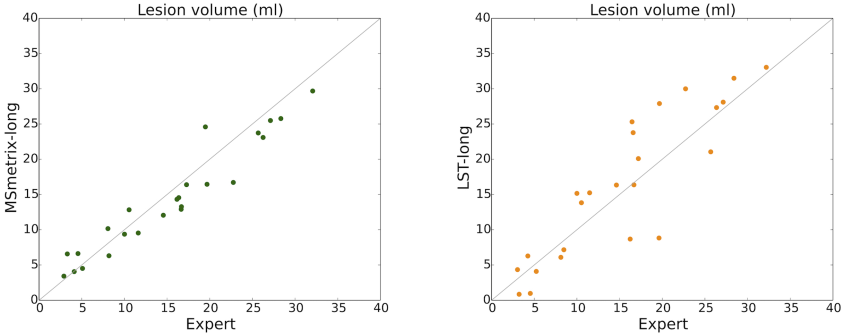


Figure 4.5: Scatter plot of total lesion volume (ml) for reference expert segmentation versus (a) MS**metrix**-long and (b) LST-long.

Table 4.1: Quantitative metrics (voxel-by-voxel, lesion and volumetric level) for measuring the cross-sectional accuracy of the automatic methods MS**metrix**-long, MS**metrix**-cross and LST-long with respect to expert segmentations on dataset 1 (n = 24).

Automatic method	Dice	F1 score	LTPR	LFPR	Absolute volume difference (ml)	r value
MSmetrix-long	0.63 (0.49 - 0.68)	0.61 (0.54 - 0.63)	0.50 (0.43 - 0.59)	0.25 (0.20 - 0.37)	2.09 (1.77 - 3.18)	0.96
MSmetrix-cross	0.60 (0.46 - 0.66)**	0.56 (0.52 - 0.61)*	0.57 (0.52 - 0.65)**	0.48 (0.36 - 0.55)**	1.48 (0.81 - 2.59)	0.95
LST-long	0.60 (0.47 - 0.65)*	0.48 (0.37 - 0.53)**	0.42 (0.30 - 0.52)*	0.40 (0.30 - 0.47)	2.66 (1.52 - 4.84)*	0.88

Except r value, all metrics are reported in median (first quartile - third quartile).
LTPR=lesion-wise true positive rate, LFPR=lesion-wise false positive rate, r value=Pearson correlation coefficient.
* Values significantly different from MS**metrix**-long (paired t-test with $p < 0.05$ significance level).
** Values significantly different from MS**metrix**-long (paired t-test with $p < 0.01$ significance level).

Table 4.2 summarises the lesion-wise change accuracy performance of MS**metrix**-cross, MS**metrix**-long and LST-long on dataset 1 in a quantitative way. In case of new lesions, MS**metrix**-long has improved over MS**metrix**-cross in the median F1 score and LFPR. Compared to LST-long, MS**metrix**-long has a higher median F1 score and LTPR. In case of enlarging lesions, MS**metrix**-long has improved over MS**metrix**-cross in the median LFPR, with marginally better F1 score. Compared to LST-long, MS**metrix**-long has a higher median F1 score, LTPR and LFPR. When new and enlarging lesions are combined, MS**metrix**-long has better correlation (r = 0.77) with the expert segmentations compared to MS**metrix**-cross (r = 0.63) and LST-long (r = 0.53). In case of absolute lesion-wise change difference, MS**metrix**-long has marginally better performance over MS**metrix**-cross and LST-long, however, with better correlation with the lesion-wise change difference of the expert segmentations (r = 0.84) compared to MS**metrix**-cross (r = 0.65) and LST-long (r = 0.72).

Table 4.2: Lesion-wise quantitative metrics for measuring the lesion change accuracy of the automatic methods MS**metrix**-long, MS**metrix**-cross and LST-long with respect to expert lesion segmentations changes on dataset 1.

	New lesions			**Enlarging lesions**			**New & enlarging lesions**
	F1 score	**LTPR**	**LFPR**	**F1 score**	**LTPR**	**LFPR**	**r value**
MSmetrix-long	0.42 (0 - 0.55)	0.33 (0 - 0.60)	0 (0 - 0.38)	0.69 (0.56 - 0.81)	0.62 (0.53 - 0.69)	0.16 (0 - 0.51)	0.77
MSmetrix-cross	0.20 (0.0 - 0.62)	0.33 (0 - 0.52)	0.50 (0.31 - 0.75)	0.68 (0.58 - 0.80)	0.59 (0.53 - 0.69)	0.24 (0.15 - 0.43)	0.63
LST-long	0 (0 - 0.43)	0 (0 - 0.29)	0 (0 - 0)	0.60 (0.51 - 0.69)	0.50 (0.35 - 0.60)	0.33 (0.15 - 0.51)	0.53

	Absolute lesion-wise change difference	
		r value
MSmetrix-long	1 (1 - 3.5)	0.84
MSmetrix-cross	1.5 (1 - 3.75)	0.65
LST-long	2 (1 - 3.5)	0.72

Except r value, all metrics are reported in median (first quartile - third quartile).
LTPR=lesion-wise true positive rate, LFPR=lesion-wise false positive rate, r value=Pearson correlation coefficient.
Here, the t-test is not performed, as the sample size is small (n = 12).

Table 4.3: The Dice score, the number (Nr.) of new lesions and the absolute volume difference (Abs. vol. diff.) between both time points for measuring the accuracy of the automatic methods MS**metrix**-long, MS**metrix**-cross and LST-long on dataset 2.

	Dice	**Nr. of new les**	**Abs. vol. diff. (ml)**
MSmetrix-long	0.89 (0.85 - 0.91)	0 (0 - 1)	0.11 (0.03 - 0.32)
MSmetrix-cross	0.69 (0.56 - 0.73)**	3.5 (1 - 5)**	0.3 (0.17 - 0.54)*
LST-long	1 (1 - 1)**	0 (0 - 0)**	0 (0 - 0.01)**

All metrics are reported in median (first quartile - third quartile).
* Values significantly different from MS**metrix**-long (paired t-test with $p < 0.05$ significance level).
** Values significantly different from MS**metrix**-long (paired t-test with $p < 0.01$ significance level).

4.3.2 Reproducibility results on dataset 2

Figure 4.6 shows an example of lesion segmentation obtained by MS**metrix**-cross, MS**metrix**-long and LST-long on a patient from dataset 2 (n = 60). Both MS**metrix**-long and LST-long are more consistent in lesion segmentation compared to MS**metrix**-cross. Compared to LST-long, MS**metrix**-long also shows better reproducibility in segmenting small lesions. Quantitatively, LST-long has the best median Dice with zero error in detecting new lesions and absolute volume difference between both time points. MS**metrix**-long has improved in the median Dice, with median error in detecting new lesions and absolute volume difference over MS**metrix**-cross (see Table 4.3). The reproducibility of LST-long is highest because it segments the most certain hyper-intense lesions in both time points at the expense of missing substantial amount of less hyper-intense lesions as shown in figure 4.6.

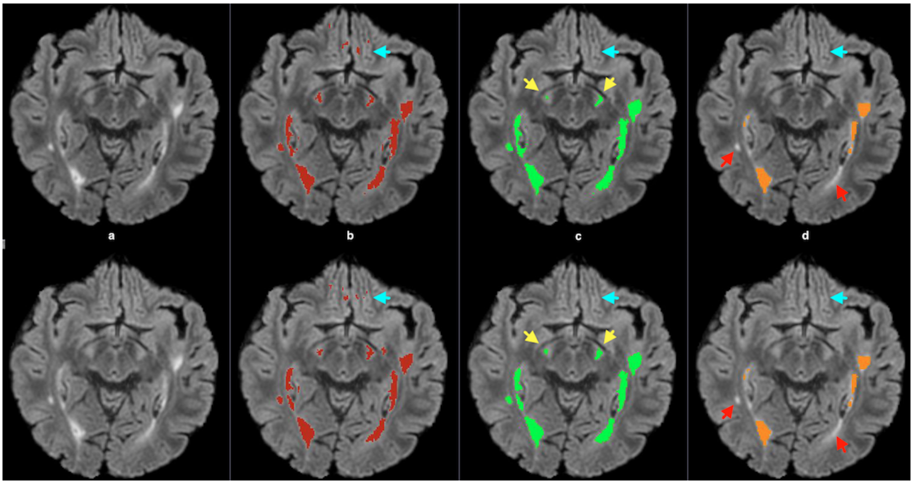


Figure 4.6: Bias corrected FLAIR image (a) followed by super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) MS**metrix**-long, and (d) LST-long. The first row corresponds to the lesion segmentation of time point 1 and the second row corresponds to the lesion segmentation of time point 2. Cyan arrows show some false positives in MS**metrix**-cross, which are absent in MS**metrix**-long. Yellow arrows specify places where MS**metrix**-long has consistently segmented some small lesions and red arrows indicate regions where LST-long misses some potential lesions.

4.4 Discussion and conclusions

Accurate and consistent lesion segmentation is very important in monitoring the MS disease progression. As manual lesion segmentation is time consuming and suffers from inter- and intra-rater variability, automated methods have the advantage of being fast and consistent. The vast majority of automatic methods are cross-sectional in nature and the average accuracy (Dice) of these methods is sufficiently high, however, these cross-sectional methods seldom report results on the lesion evolution accuracy and this hinders a fair comparison of our method against them. Moreover, another factor to consider is whether the segmentation method is supervised or unsupervised. We compare our unsupervised method with other unsupervised methods only because supervised methods often require a representative training dataset, including expert segmentation, in order to build a model that can be used on new patients for lesion segmentation. This training dataset is very difficult to build because MS lesions have all possible shapes, intensities and are heterogeneously distributed in the WM. Moreover, the new image to be segmented should be well represented in the training dataset which is not always possible. Two well known publicly available unsupervised MS lesion segmentation tools are Lesion-TOADS [146] and LST.

We choose LST because of two reasons: (1) in chapter 2, we have shown that our cross-sectional method (MS**metrix**-cross) had a better performance compared to Lesion-TOADS in terms of accuracy and reproducibility. Since in this chapter we also report results from our cross-sectional method, we decided that the comparison with Lesion-TOADS is not required, (2) only LST tool has a longitudinal MS lesion segmentation pipeline. Thus it is logical to compare MS**metrix**-long with LST-long as both methods are unsupervised and longitudinal in nature.

In this chapter, MS**metrix**-long pipeline combines both spatial and temporal relationships of lesions for accurate and consistent lesion segmentation. The spatial relationship is based on Markov Random Field and is incorporated in MS**metrix**-cross. The temporal relationship is modelled in a joint lesion segmentation, which uses difference image and cross-sectional lesion segmentations of two time points. The difference image models the growth and shrinkage of lesions and thus helps in recovering those lesions that are missed by the cross-sectional lesion segmentation. In addition, if a lesion is present in both time points but has been segmented in only one of the time point, then the joint lesion segmentation facilitates the recovery of that lesion at the other time point. Moreover, brain atrophy has also minimal impact on the performance of MS**metrix**-long because (1) atrophy is generally small and global in nature (2) it occurs near the CSF boundary and these transitions *i.e.,* (CSF → GM and CSF → WM) are excluded in the difference image GMM model, (3) we tested global non-rigid registration in addition to affine registration *i.e.,* non-rigid registration only on a coarse level, to accommodate for the atrophy and we found out that it has a minimal, but potentially negative impact on the final lesion segmentation. Therefore, to gain computational efficiency we excluded this global non-rigid registration from MS**metrix**-long pipeline. Furthermore, we made a choice of using two time points instead of multiple time points for lesion segmentation as it is unclear whether clinical users would benefit from a long-term longitudinal analysis method. Indeed, incorporating multiple time points at once would be more consistent and more robust as more information is available compared to two time points. On the other hand, longitudinal consistency over multiple time points might perhaps also influence the accuracy of individual lesion segmentation. Change in the imaging protocols in a longitudinal follow-up also makes multiple time point analysis more difficult. Moreover, radiological reporting guidelines recommend comparison of current MRI against previous time point [169]. Lesion changes between successive time points (e.g., new lesions) are relevant information for clinicians, as it is essential to capture disease evolution, instead of disease history. Thus, a two-time point approach can always be applied between successive visits, without the need to incorporate a longer history into the analysis. Once the method is feasible for two time points, it can be applied in a multi-time point fashion. We have actually performed a study on a large retrospective dataset from a clinical trial (unfortunately, we are not allowed to publish the results in this thesis), where patients were annually monitored for a longer period of time. When grouping time points in pairs, *e.g.,* (TP1, TP2) and (TP2, TP3), we observed a high consistency in total lesion volume at the middle time point (TP2), without having to use more than 2 time points simultaneously.

Among the methods proposed in the literature for longitudinal lesion segmentation, our approach has some similarities to [48] and [2], which are also based on EM frameworks. In contrast with [48], our method is unsupervised and can segment new, enlarging, disappearing and shrinking lesions. As opposed to [2], our joint EM model takes cross-sectional lesion segmentation as prior information on the lesion class in both time points and processes each time point in its own space to avoid bias in the lesion segmentation.

In order to evaluate the effect of the pruning step, we also calculated the cross-sectional accuracy (Dice, LTPR and LFPR) of MS**metrix**-long after the joint lesion segmentation step. The Dice, LTPR and LFPR (reported in median (first quartile - third quartile)) after the joint lesion segmentation step are 0.60 (0.45 - 0.65), 0.64 (0.54 - 0.69) and 0.81 (0.72 - 0.87) respectively. Comparing these results with the the voxel-by-voxel accuracy of MS**metrix**-long after the pruning step (see table 4.1), we observe that the pruning step increases the overall Dice score by decreasing the false positive rate at the expense of a decrease in true positive rate.

In order to investigate the cause of low lesion-wise true positive rate for cross-sectional accuracy of MS**metrix**-long compared to MS**metrix**-cross (see table 4.1), we calculated the average lesion-wise true positive rate for small (0.003 ml - 0.01 ml), medium (0.01 ml - 0.05 ml) and large ($>$ 0.05 ml) lesion volumes. The average lesion-wise true positive rate for MS**metrix**-long and MS**metrix**-cross for small lesions is 0.13 and 0.27 respectively, followed by medium lesions 0.30 and 0.37 and large lesions 0.75 and 0.81. It can be seen that MS**metrix**-long misses more small and medium size lesions. The primary cause of missing these lesions is that they are either iso-intense with grey matter intensities (thus missed by intensity threshold mask used in the pruning step) or they are removed by the binary false positive mask (used in the pruning step). However, it is important to note that both intensity threshold mask and binary false positive mask play a key role in reducing the false positives as described in the previous paragraph.

One important aspect of MS**metrix**-long is that its performance is dependent on the cross-sectional lesion segmentation. This suggests that if MS**metrix**-cross has either consistently missed a lesion, or segmented a non-lesion at both time points, then it will be either missed or retained by MS**metrix**-long, respectively. As presented in the result section, MS**metrix**-long is more accurate and reproducible than MS**metrix**-cross. The increase in cross-sectional accuracy (Dice, F1 score) and lesion change accuracy for new lesions (F1 score) is due to the reduction in lesion-wise false positive rate using the lesion segmentation information from the other time point. For enlarging lesions, a marginal increase in the median F1 score is observed for MS**metrix**-long due to larger differences in the lesion segmentation boundary between the expert and MS**metrix**-long. MS**metrix**-long has also slightly better absolute lesion-wise change difference compared to MS**metrix**-cross primarily due to a reduction in lesion-wise false positive rate. A modest decrease in the absolute volume difference is due to the under-segmentation of lesions by MS**metrix**-long (figure 4.5) and the elimination of a few lesions that are close to the cerebral cortex. Interestingly, a substantial lesion-wise false positive rate in MS**metrix**-cross suggests that the false lesions compensate towards missed lesions volume resulting in a lower absolute volume difference compared to MS**metrix**-long. The significant improvement in reproducibility (Dice, number of new lesions and absolute volume difference) of MS**metrix**-long could also be explained by the benefit of using the lesion segmentation of the other time point.

In comparison to LST-long, MS**metrix**-long is more accurate (Dice, F1 score) and slightly less reproducible. Cross-sectionally, LST-long has higher absolute volume difference and lesion-wise false positive rate; lower lesion-wise true positive rate and F1 score on dataset 1. The high absolute volume difference of LST-long could be explained by the over-segmentation of lesion boundaries. A high lesion-wise false positive rate of LST-long could be explained by the segmentation of FLAIR artefacts or cortical foldings as lesions. For the lesion change accuracy, MS**metrix**-long has superior performance for all measures compared to LST-long. This could be explained by the fact that LST-long segments the most hyper-intense lesions and is thus very consistent (see table-4.3), but misses many small less hyper-intense lesions (figure 4.4, figure 4.6).

In conclusion, we have presented MS**metrix**-long: an iterative two time point white matter lesion segmentation method based on a joint EM framework using two time points. The proposed method is unsupervised and can segment new, enlarging, disappearing, shrinking and static lesions. We first analyse both time points separately followed by a joint lesion segmentation, which models the lesion evolution as a Gaussian mixture model. The accuracy and reproducibility of MS**metrix**-long is compared with MS**metrix**-cross and the publicly available lesion segmentation tool LST-long on two datasets that are representative for clinically feasible acquisition protocols. MS**metrix**-long has outperformed MS**metrix**-cross. Compared to LST-long, MS**metrix**-long has better accuracy and similar reproducibility.

Chapter 5

Clinical relevance of new and enlarging lesion volume in RRMS

Abstract

MRI has become an integral part in monitoring multiple sclerosis disease. Several MRI markers have been explored in this direction, *e.g.,* total lesion load and new and enlarging lesions count. However, the volume of the new and enlarging lesions (NEL) as an MRI marker has never been explored for its possible relationship with the most widely used clinical marker: Expanded Disability Status Scale (EDSS) score. In this study, we explored the association of these two markers on RRMS patients. NEL volume is computed using MS**metrix**-long, where only the enlarged part of the enlarging lesion is used for the volume computation together with the new lesion volume. The relationship between NEL volume and EDSS score is explored both at group level and on an individual level using a multi-centre dataset consisting of images from three centres. NEL volume correlates better with EDSS score's evolution at individual level than at a group level, where a low correlation is observed. Nevertheless, this study shows the applicability of our method on multi-centre datasets without re-tuning or re-training, which makes it useful for the clinical use.

This chapter has been accepted for poster presentation at the conference: *The consortium of Multiple Sclerosis Centres* (2017), Jain, S., Sima, D.M., Van Vlierberghe, E., Dubois, B., Dupont, P., Kocevar, G., Durand-Dubief, F., Sappey-Marinier, D., Wang, C., Barnett, M.H., Van Huffel, S., Maes, F. and Smeets, D. Clinical relevance of new and enlarging lesion volume in Relapsing Remitting Multiple Sclerosis: A multi-centre study.

5.1 Introduction

MRI of the brain has shown its effectiveness in MS disease in various dimensions such as disease diagnosis [119, 148], monitoring, prognosis [11, 114, 56] and testing of drug efficacy [105, 157]. Towards this end, several MRI markers have been explored such as total lesion volume [51, 53, 22], T1 hypointense lesions (black holes) [10, 110, 106, 68], whole brain atrophy [120, 55, 147, 81] and new lesions (based on contrast enhanced T1-weighted image and/or T2-weighted/FLAIR image) count and volume [33, 44, 3]. Total lesion volume provides information on the cumulative disease burden. Black holes are seen as a marker of neuronal destruction and axonal loss, and have shown correlation with EDSS score [178]. Whole brain atrophy provides insights on the brain tissue loss and is associated with the long term cognitive impairment, fatigue and increase in the EDSS score [39, 156]. New lesions based on the T2 -eighted/FLAIR image suggest an inflammatory reaction that has developed between two MRI scans and provide information on the disease activity and disease progression. New lesions based on the contrast enhanced T1-weighted image are associated with the active inflammation in the brain due to the disruption of the blood-brain barrier and have shown to be more correlated than T2/FLAIR lesions with the number of relapses (attack when MS symptoms flare up) that a patient has experienced [3, 44]. It is important to mention that the correlation of MRI markers with clinical markers is not very strong [157, 155, 95], as MRI is more sensitive in detecting disease activity earlier in time [119].

Enlarging lesions as a potential MRI marker have also been investigated for monitoring the disease activity and testing of drug efficacy [3, 155, 133]. There are not many studies that include enlarging lesions. A possible reason could be that these studies generally have an expert panel, which is required to examine every enlarging lesion segmentation in longitudinal scans, and this is very time consuming. Automatic MS lesion segmentation tools can be used to find enlarging lesions; however, counting the enlarging lesions using these softwares is often debatable.

The aim of this study is to investigate the role of new and enlarging lesion volume in monitoring the RRMS disease activity using MS**metrix**-long, as described in chapter 4. We choose this because to the best of our knowledge, volume of the new and enlarging lesions has never been explored for its association with the EDSS score in RRMS patients. In particular, we explore the relationship between the new and enlarging lesion volume and change in the EDSS score, both on a group level and on an individual patient level using multi-centre datasets.

5.2 Methods

The new and enlarging lesion volume between two consecutive time points is computed using the lesion segmentations obtained from MS**metrix**-long. A new lesion is defined as the lesion present only at the second time point and an enlarging lesion is defined as the lesion whose overall volume is increased by more than 5% compared to the previous time point. In this study we used the volume of the enlarged part of the enlarging lesion, which defines the increment in the lesion volume.

5.2.1 Data

Dataset 1 contains scans from 30 RRMS patients acquired on a SIEMENS 1.5T scanner (Sonata), each scanned at least three times at an interval of approximately 6 months or 12 months. Each time point contained two sequences: 3D T1-weighted and 2D FLAIR, and their details are described in table 5.1. For this dataset the EDSS score was available at all MRI examination dates. This study was carried out in accordance with the recommendations of the French national agency for medicine and health products safety (ANSM) and was approved by the local ethical committee (CPP Sud-Est IV). All subjects gave written informed consent.

Dataset 2 contains scans from 20 RRMS patients acquired on a GE 3T scanner (Discovery MR750), each scanned twice at an interval of approximately one year. Each time point contained two 3D sequences: T1-weighted and FLAIR, and their details are described in table 5.1. For this dataset the EDSS score was available at both MRI examination dates along with one in the future time (no MRI scan for the future time). This study was reviewed and conducted within the guidelines set out in the National Statement on Ethical Conduct in Human Research (2007) in Australia, and approved by University of Sydney Human Research Ethics Committee. All subjects gave written informed consent.

Dataset 3 contains scans from 35 RRMS patients acquired on a Philips 3T scanner, each scanned at least twice at a variable time interval ranging from approximately 4 to 12 months. Each time point contained two 3D sequences: T1-weighted and FLAIR, and their details are described in table 5.1. For this dataset the EDSS scores were not always available nearby the MRI examination date, therefore we used only those MRI scans where at least two EDSS scores were available having a maximum delay of ± 6 months from their closest MRI examination date. 4 out of 35 subjects had matching EDSS scores and MRI scans available and only these 4 subjects were used in this study. This study was carried out in accordance with the recommendations of the Belgian and

Table 5.1: Sequences description for all three datasets.

Sequence	TR (ms)	TE (ms)	TI (ms)	FOV (mm^2)	Voxel size (mm^3)	No. of slices
			Dataset 1			
3D T1-weighted MPRAGE	1970	3.93	1100	256 × 256	1 × 1 × 1	176
2D FLAIR TSE	8000	105	2200	180 × 240	0.94 × 0.94 × 3	46
			Dataset 2			
3D T1-weighted IR-FSPGR	7.2	450	2.8	230 × 230	0.9 × 0.9 × 0.9	188
3D FLAIR CUBE	8000	165	2179	240 × 240	0.60 × 0.47 × 0.47	280
			Dataset 3			
3D T1-weighted MPRAGE	9.64	4.6	NA	250 × 250	0.98 × 0.98 × 1.2	140
3D FLAIR SENSE	4800	360.2	1650	250 × 250	1 × 1 × 0.55	321

NA = Not available.

Dutch legislation and was approved by the UZ Leuven ethical committee. All subjects gave written informed consent.

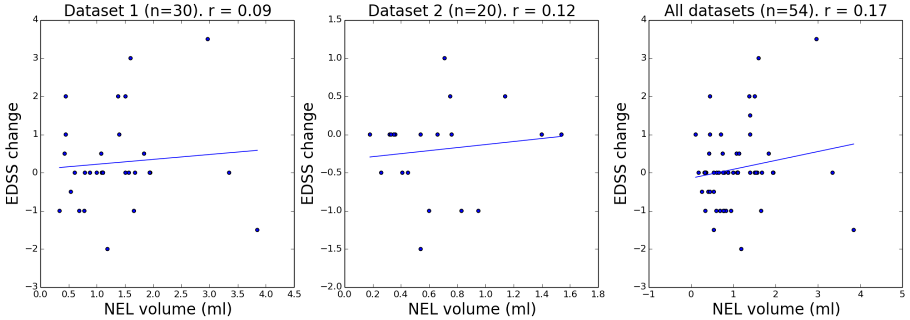
5.2.2 Study design

We analyse the relationship between new and enlarging lesion volume and EDSS score both at group level and individual level. At group level, the correlation between EDSS change (difference between the last and the first EDSS score) and new and enlarging lesion volume is investigated using Pearson correlation coefficient (r value). Since dataset 3 has only four subjects that had matching EDSS scores and MRI scans, no separate group analysis was performed for this dataset. At individual level, for every subject, EDSS scores and new and enlarging lesion volumes are plotted against their corresponding examination dates as time points. The aim of the individual study is to analyse whether the evolution of new and enlarging lesion volume can correlate with the evolution of the EDSS score. Since dataset 2 has only two time points, it is excluded from the individual study.

5.3 Results

5.3.1 Group analysis

Figure 5.1 shows the scatter plots between EDSS change and new and enlarging lesion volume for dataset 1, dataset 2 and all datasets combined. Although the correlation is low, it increases when the values from all three datasets are included simultaneously in the analysis.


Dataset 1 (n=30). r = 0.09

Dataset 2 (n=20). r = 0.12

All datasets (n=54). r = 0.17

Figure 5.1: Scatter plots between EDSS change and new and enlarging lesion volume for dataset 1 (left), dataset 2 (middle) and all datasets combined (right).

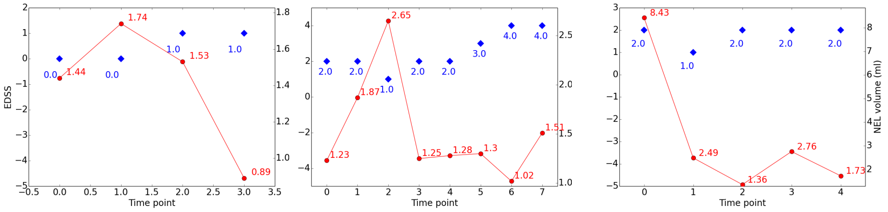
5.3.2 Individual analysis

Dataset 1

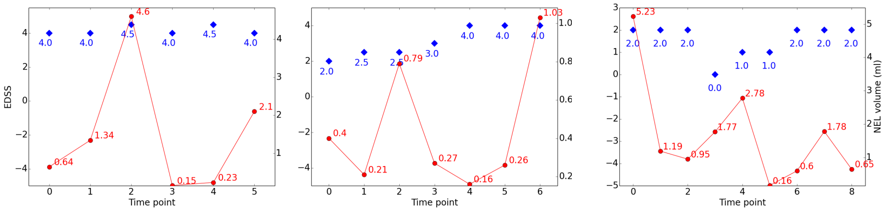
Figures 5.2, 5.3 and 5.4, show the evolution of EDSS score and new and enlarging lesion volume for each subject over time for dataset 1. From the figures 5.2 and 5.3 (n = 21) it can be seen that with an increase in the new and enlarging lesion volume, the EDSS score is increased either at the current time point or in the future time point. With a decrease in the new and enlarging lesion volume, the EDSS score is either decreased or it remains nearly the same. Lastly, when the new and enlarging lesion volume equally increases and decreases over time (random behaviour), the EDSS score fairly remains constant. However, the remaining subjects (n = 9) in the figure 5.4 exemplify the cases where this trend is absent.

Dataset 3

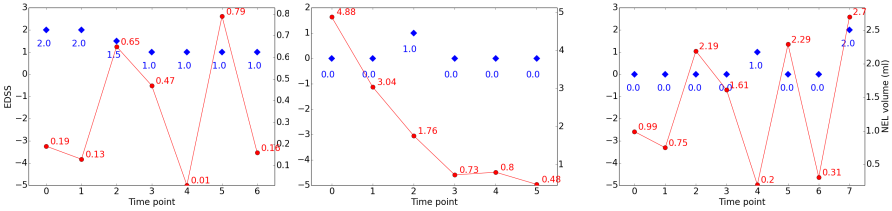
Figure 5.5 shows the evolution of EDSS score and new and enlarging lesion volume for each subject over time for dataset 3. Three out of four subjects show the same trend as observed in dataset 1 (with an increase in the new and enlarging lesion volume, the EDSS score is increased either at the current time point or in the future. With a decrease in the new and enlarging lesion volume, the EDSS score is either decreased or it remains nearly the same). An opposite trend is observed for the subject 4.



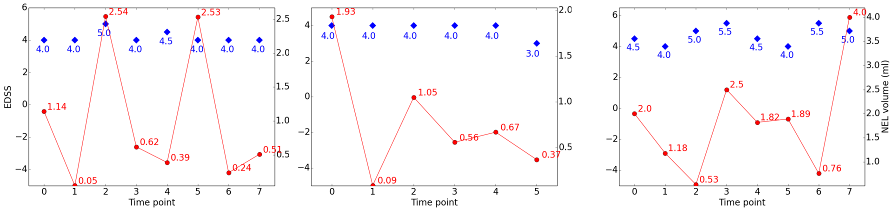
(a) Subject 1 (b) Subject 2 (c) Subject 3



(d) Subject 4 (e) Subject 5 (f) Subject 6

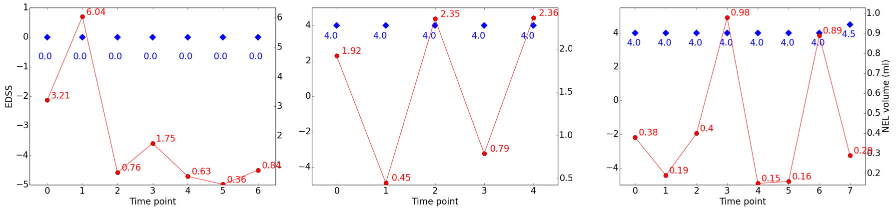


(g) Subject 7 (h) Subject 8 (i) Subject 9

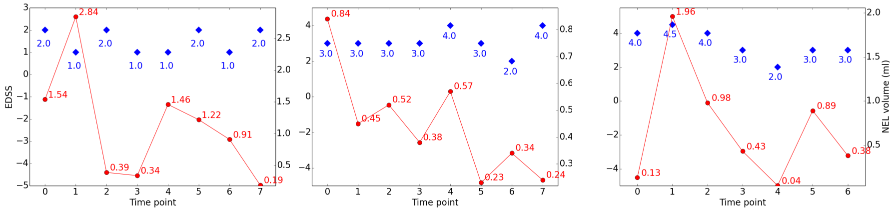


(j) Subject 10 (k) Subject 11 (l) Subject 12

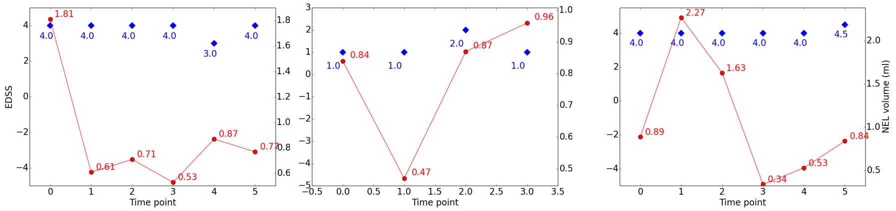
Figure 5.2: Examination time points versus EDSS score (in blue) and new and enlarging lesion volume (in red) for dataset 1.



(a) Subject 13 (b) Subject 14 (c) Subject 15



(d) Subject 16 (e) Subject 17 (f) Subject 18

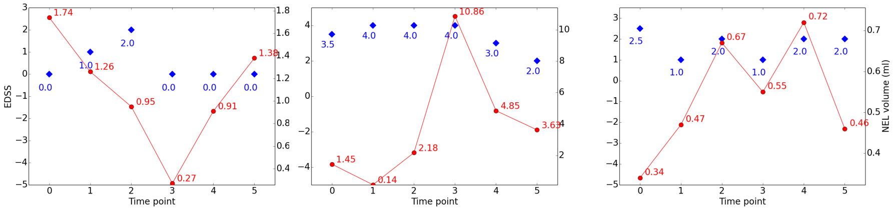


(g) Subject 19 (h) Subject 20 (i) Subject 21

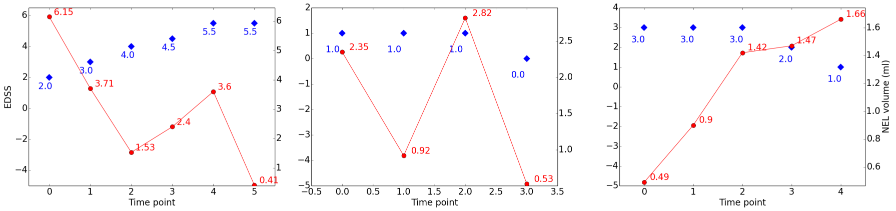
Figure 5.3: Examination time points versus EDSS scores (in blue) and new and enlarging lesion volumes (in red) for dataset 1.

5.4 Discussion and conclusions

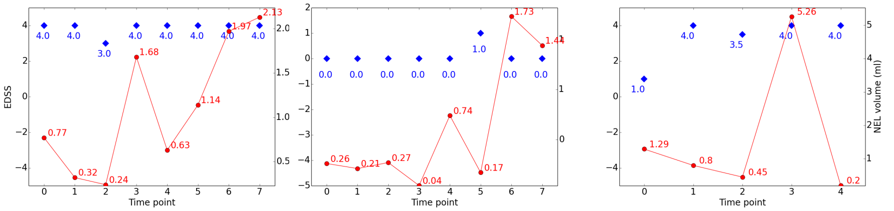
Accurate and reliable quantification of MRI derived measures such as new and enlarging lesion volume is very important. Majority of the literature studies quantify these measures using either expert segmentations or using automated segmentation tools. While the former is most accurate, it often suffers from intra and inter rater variability, and has low reproducibility of lesion segmentations on the follow-up time points. Compared to expert lesion segmentations, automated lesion segmentation tools are generally more reproducible but less accurate. Additionally, large majority of MS lesion segmentation tools are cross-sectional (based on a single time point) in nature, which affects their reproducibility. Automated longitudinal MS lesion segmentation tools address this problem and optimise for accuracy and reproducibility simultaneously using using lesion segmentation information from two or more time points. In this study we have



(a) Subject 22 (b) Subject 23 (c) Subject 24



(d) Subject 25 (e) Subject 26 (f) Subject 27

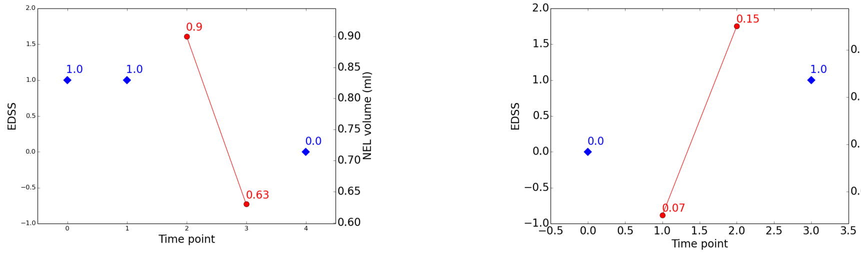


(g) Subject 28 (h) Subject 29 (i) Subject 30

Figure 5.4: Examination time points versus EDSS scores (in blue) and new and enlarging lesion volumes (in red) for dataset 1.

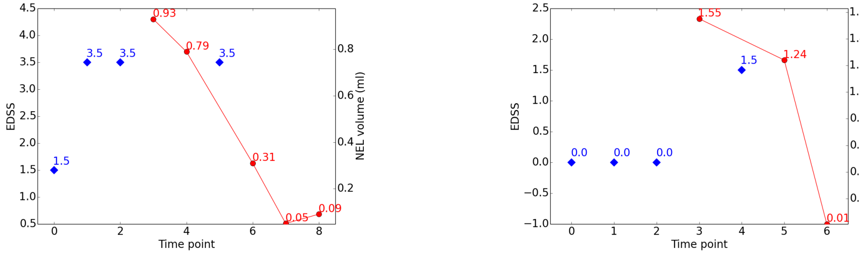
chosen a two time points MS lesion segmentation tool (MS**metrix**-long) that has shown to have a good balance between accuracy and reproducibility (see chapter 4).

Quantification of new and enlarging lesion volume from the obtained lesion segmentations is also challenging. The two main problems are: (a) lesion boundary error between two time points causing a lesion to be classified as an enlarging lesion or not (b) choice of enlarging lesions *i.e.,* considering lesions that are only enlarged (*i.e.,* no shrinking) or also consider lesions that have enlarged at one location and shrunk at another location. Since we use MS**metrix**-long, which is more reproducible, the severity of the first problem is reduced. For the second problem, we assumed that the enlarging part of the enlarging lesion (irrespective of whether it has shrunk or not) is more important for disease monitoring since it indicates active inflammation.



(a) Subject 1

(b) Subject 2



(c) Subject 3

(d) Subject 4

Figure 5.5: Examination time points versus EDSS scores (in blue diamond) and new and enlarging lesion volumes (in red circle) for dataset 3.

Another challenge in our study was the availability of a dataset that satisfies our requirements. It includes: (1) availability of EDSS score near MRI examination date (2) periodic follow up of patients (3) no therapy change. As the EDSS score is the patient's disability state at a given time point, it is necessary that the MRI scan was taken nearby *e.g.,* ± 6 months. Periodic follow-up of patients helps in analysing the lesion evolution pattern and investigating its relationship with EDSS evolution over time. In this study, the increase in new and enlarging lesion volume seems to show an increase in the EDSS score in a near future. However, as the number of subjects used in this study is low, a larger population study is needed to confirm this observation. Lastly, it is also important that the patient' therapy is not changed over time as it will influence both MRI and clinical measurements. All these requirements are very difficult to meet in clinical practice and thus pose a big challenge.

In this study we explored the relationship between EDSS score and new and enlarging lesion volume using three different multi-centre datasets. The new and enlarging lesion volume was computed with MS**metrix**-long and the analysis was performed both on a group and on an individual level. On a group level, a weak correlation between EDSS score and new and enlarging lesion volume could be explained by the scientific evidence that MRI is a more sensitive disease activity marker than the EDSS score [119]. We also observed a similar trend in our individual study where the change in new and enlarging lesion volume happened before the actual change in the EDSS score. Note that the new and enlarging lesion volume may also relate to other symptoms, such as cognitive scores (which are not captured in the EDSS score), and will be explored in the future.

In conclusion, this study showed interesting trends in the data. In majority of the cases, NEL volume correlates better with EDSS score's evolution at individual level, although at a group level, the correlation between these two variables is low. An extensive further validation is required to explore other confounding factors. Nevertheless, this study shows the applicability of our method on multi-centre datasets without re-tuning or re-training, which makes it useful for the clinical use.

Chapter 6

Conclusions and future work

6.1 Main contributions

Accurate and consistent MS lesions segmentation has been demonstrated to be a valuable step in the clinical routine evaluation of MS patients. Enhancing the spatial resolution of quantified MR spectroscopic maps in MS patients can help for better characterisation of small lesions.

6.1.1 Methods

In this work, the following scientific contributions were made:

- *MS**metrix**-cross*: An iterative, EM based method that segments a MR image of an MS brain into GM, WM, CSF and lesions. A GMM model is used to model tissue class intensities of GM, WM and CSF and lesions are detected as an outlier to this GMM model.
- *PBSR*: This super-resolution technique uses MS**metrix**-cross outputs as prior knowledge to regularise the reconstruction process. The algorithm iteratively estimates the expected metabolite value in every high-resolution MRSI voxel using segmentations and tissue intensities information; and then corrects these expected metabolite values using the input low resolution MRSI map.

- *MS**metrix**-long*: An iterative, joint EM based method for two time points lesion segmentation. The method uses MS**metrix**-cross outputs of the two time points as prior knowledge along with the FLAIR based difference image to improve the lesion segmentation at both time points.

6.1.2 Results and discussion

The results presented in this thesis addressed two main problems: MS lesion segmentation and MRSI super-resolution.

MS lesion segmentation

MSmetrix-cross
We showed that MS**metrix**-cross has good accuracy and reproducibility with some systematic undersegmentation and it misses small lesions that are less than five voxels. Comparing with state-of-the-art methods, the accuracy of MS**metrix**-cross outperformed two lesion segmentation tools: LST and Lesion-TOADS. Regarding the reproducibility, MS**metrix**-cross and LST had similar performance and both methods outperformed Lesion-TOADS.

MSmetrix-long
We showed that MS**metrix**-long consistently segments lesions at two time points by decreasing the lesion-wise false positive rate with slight reduction in lesion-wise true positive rate. The method's performance is dependent on the MS**metrix**-cross and thus requires good cross-sectional lesion segmentation. Comparing with state-of-the-art methods, the accuracy and reproducibility of MS**metrix**-long outperformed MS**metrix**-cross. In comparison to LST-long, MSmetrix-long is more accurate and slightly less reproducible (LST-long segments the most hyper-intense lesions). Moreover, initial results showed that new and enlarging lesion volume may correlate with the EDSS score's evolution for individual patients. However, a low correlation between these two variables at group level requires an extensive validation to explore other confounding factors.

MRSI super-resolution

We showed that the patch-based super-resolution method preserves tissue contrast and structural information compared to conventional interpolation methods, and matches well with the trend of acquired high-resolution MRSI. The method cannot recover small scale features that were not sufficiently expressed in the low resolution MRS image and is applicable only if the difference between the image resolution of MRI and MRSI images is less than a scale factor of four.

6.2 Future perspectives

We believe that both patch based super-resolution and two time points lesion segmentation methods can be further improved, and we now mention some possibilities in that direction.

Patch-based super-resolution

In chapter 3, the PBSR method uses a parameter α that controls the contribution of MRI and MRSI images in the super-resolution process, and was derived using lesion segmentation of FLAIR image. As the actual damage is beyond what is seen on the FLAIR image, MRSI should have more contribution in the vicinity of the lesion. However, the important question is, to what extent? For this, advanced imaging methods such as magnetic transfer imaging and diffusion tensor imaging that are more sensitive than conventional MRI and have better resolution than MRSI could be used to derive this information. However, care should be taken that the resulting high-resolution image is plausible and does not create features that are not seen in MRSI. Also, the benefit of using multi-scale resolution approach could not be tested on the real images as the data was already acquired and thus the resolution was fixed. With the use of a 7T scanner and parallel imaging techniques, we hope to acquire MRSI datasets from more patients at 2×2 mm^2, 3×3 mm^2 and 4×4 mm^2 in-plane resolution and thus test the benefit of the multi-scale approach and further validate our method.

In the present study, only NAA and myo-Ins metabolites are used as they are one of the most common metabolites of interest in MS. Conceptually, the super-resolution method developed could easily be applied to upsample any metabolite of interest. Simultaneous multiple metabolites upsampling is also very interesting to investigate, as the information from each metabolite aids in the reconstruction of other metabolites.

Two time points lesion segmentation

In this thesis, two time points lesion segmentation currently segments MS lesions of the two time points. However, temporal consistency of the other tissue classes can also be incorporated in the same framework using a temporal Markov Random Field. One of the important challenges in developing this would be computation time, which is a crucial factor in real time. Using the graphical processing unit (GPU) can alleviate this problem to a great extent. Additionally, histogram normalisation is based on cumulative histogram matching, which may be sensitive to lesion changes as these would show up in the histogram, *e.g.,* normalising an image with low lesion load with an image with significant lesion load may result in histograms that are representing two scenarios. This problem could be addressed by doing tissue based intensity normalisation such as proposed in [82], where the joint histograms (*e.g.,* T1-weighted and FLAIR) of the two time points are non-rigidly registered to find the intensity transformation function. Finally, we observed that the lesion-wise true positive rate is decreased from our single time point to two time points lesion segmentation methods. It would be a great challenge to improve the lesion-wise true positive rate without compromising on reproducibility. 'Deep learning' algorithms have become very popular nowadays and may address this problem. Deep learning methods are training based methods that can model a high level of abstraction in the data which is not obvious to the human eye. Using the MS**metrix**-cross pipeline, the required training data may be generated.

In chapter 5 we explored the relationship of new and enlarging lesions volume with the EDSS score. EDSS score does not capture all aspects of MS such as cognitive decline or disease progression, therefore, scores such as Multiple Sclerosis Functional Composite (MSFC) [36] and Multiple Sclerosis Severity Score (MSSS) [138] would be explored for their association with MRI biomarkers. To investigate and validate potential MRI biomarker/s, a large retrospective study is desirable and should be conducted with clinical biomarkers information taken periodically. Also, this study would help in making a lesion ageing population graph that can serve as a template for disease progression and evaluation of drug efficacy. Finally, according to the revised McDonald criteria, infratentorial, periventricular or juxtacortical brain regions are considered of high importance for disease diagnosis [119]. Therefore, regional lesion classification and association between lesion location and clinical symptoms is also an interesting future perspective.

From this research we conclude that the developed multiple sclerosis lesion segmentation methods, through their robustness and automation, could bring an added value to the clinical routine evaluation of multiple sclerosis patients. Also, the patch-based high-resolution MR spectroscopic images, through its tissue contrast conservation, can offer better lesion characterisation.

Appendix A

Accuracy and reproducibility of LST and Lesion-TOADS

Overview

This appendix provides additional information and supporting figures regarding the accuracy and reproducibility for publicly available lesion segmentation tools LST and Lesion-TOADS as described in chapter 2.

A.1 Accuracy results on dataset 1

Best case for LST

Figure A.1 shows the best case for LST (Dice: 0.80, sensitivity: 0.80 and precision: 0.79). This case has Dice: 0.82, sensitivity: 0.76, precision: 0.90 for MS**metrix**-cross and Dice: 0.70, sensitivity: 0.54, precision: 0.98 for Lesion-TOADS. The higher sensitivity of LST compared to the other two methods is caused by the detection and segmentation of subtle lesions (marked by a pink arrow head); however, the lower precision of LST suggests that it introduced false lesions and it overestimated the lesion boundary (see region marked by cyan arrow heads). MS**metrix**-cross has higher Dice similarity index than LST due to lower number of false positive lesion voxels.

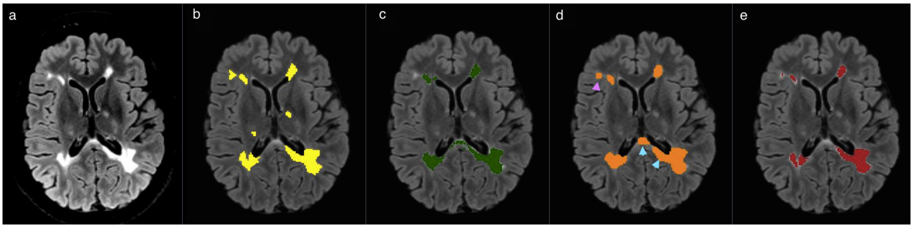


Figure A.1: Original FLAIR image (a) followed by bias corrected FLAIR image and superimposed lesion segmentation from: (b) expert segmentation, (c) MS**metrix**-cross, (d) LST, (e) Lesion-TOADS. Cyan arrow heads show false positive lesions and overestimation of the lesion boundaries in LST. Pink arrow head show lesions picked by LST but not by the other methods except partially one in Lesion-TOADS.

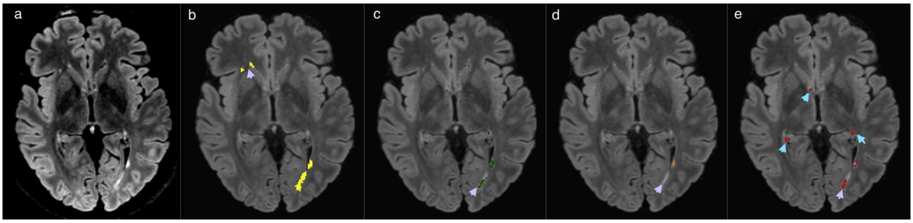


Figure A.2: Original FLAIR image (a) followed by bias corrected FLAIR image and superimposed lesion segmentation from: (b) expert segmentation, (c) MS**metrix**-cross, (d) LST, (e) Lesion-TOADS. Cyan arrow heads show false positive lesions in Lesion-TOADS. Purple arrow heads show missed subtle lesions and underestimation of lesion boundary.

Worst case for LST and Lesion-TOADS

Figure A.2 shows the worst case for LST (Dice: 0.31, sensitivity: 0.21 and precision: 0.58), which is also the worst case for Lesion-TOADS (Dice: 0.44, sensitivity: 0.38 and precision: 0.52). Here, MS**metrix**-cross has a better Dice similarity index, sensitivity and precision compared to the other two methods (Dice: 0.52, sensitivity: 0.40, precision: 0.76). The low sensitivity and precision of LST are due to the fact that it did not find the big lesion, indicated by the purple arrow head in (d). On the other hand, for MS**metrix**-cross and Lesion-TOADS, low sensitivity is due to missing subtle lesions and/or underestimation of lesion boundary (purple arrow head). Lesion-TOADS has a lower precision compared to MS**metrix**-cross because it finds a lot of false positive lesions (cyan arrow head).

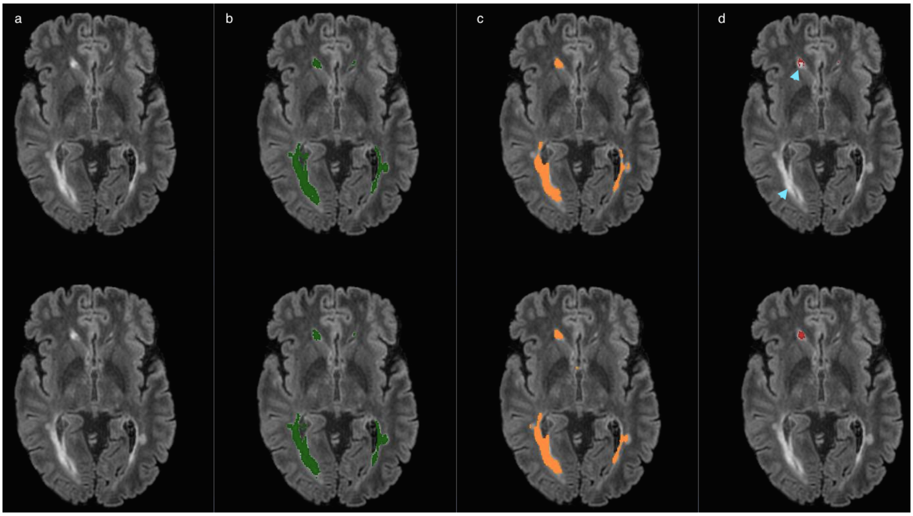


Figure A.3: Bias corrected FLAIR image (a) and super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) LST, (d) Lesion-TOADS. The first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2. Cyan arrow heads show missed lesions and difference in the lesion segmentation boundary between scan 1 and scan 2 for Lesion-TOADS.

A.2 Reproducibility results on dataset 2

Best case for LST

Figure A.3 shows the best case for LST (Dice: 0.84). In this case, MS**metrix**-cross has a similar Dice of 0.83 followed by Lesion-TOADS (Dice: 0.78). A lower Dice similarity index for Lesion-TOADS is mainly due to the inconsistent estimation of lesion boundaries (marked by cyan arrow heads) compared to MS**metrix**-cross and LST, which are more consistent in the lesion segmentation in both scans. However, on the other hand, Lesion-TOADS misses big lesions (marked by cyan arrow heads) whereas MS**metrix**-cross and LST detect them successfully.

Best case for Lesion-TOADS

Figure A.4 shows the best case for Lesion-TOADS (Dice: 0.82). In this case, LST and MS**metrix**-cross have comparable Dice of 0.79 and 0.77, respectively. The higher Dice similarity index for Lesion-TOADS compared to MS**metrix**-cross and LST is mainly due to its quite consistent performance in estimation of lesion boundaries in scan 1 and scan 2 for this case. A lower Dice similarity

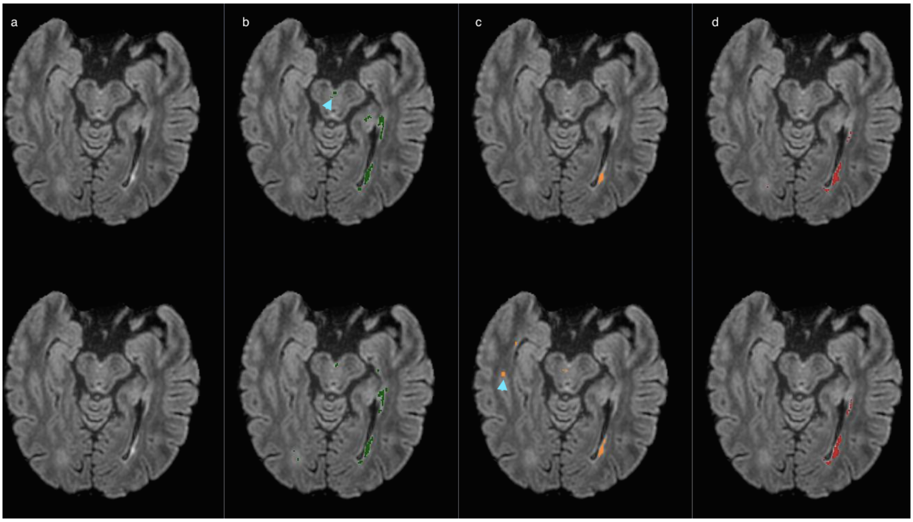


Figure A.4: Bias corrected FLAIR image (a) and super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) LST, (d) Lesion-TOADS. The first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2. Cyan arrow heads represent false lesion detection for MS**metrix**-cross and LST.

index for both MS**metrix**-cross and LST accounts for (probably) several false lesions in either of the scans (marked by cyan arrow heads).

Worst case for LST

Figure A.5 shows the worst case for LST (Dice: 0). In this case, MS**metrix**-cross and Lesion-TOADS have comparable Dice of 0.38 and 0.35, respectively. Before we explain the results, it is important to mention here that this subject has very few lesions. A zero Dice similarity index for LST is primarily due to the fact that it is unable to find any lesions in scan 1, but it finds some lesions in scan 2. Both MS**metrix**-cross and Lesion-TOADS consistently find both true lesions (pink arrow head) and false lesions (cyan arrow head) across the scans. However, the lower Dice similarity index for Lesion-TOADS accounts for slightly more false lesion detection in either of the scans compared to MS**metrix**-cross.

Worst case for Lesion-TOADS

Figure A.6 shows the worst case for Lesion-TOADS (Dice: 0.15). In this case, LST has the best performance (Dice: 0.63) followed by MS**metrix**-cross (Dice:

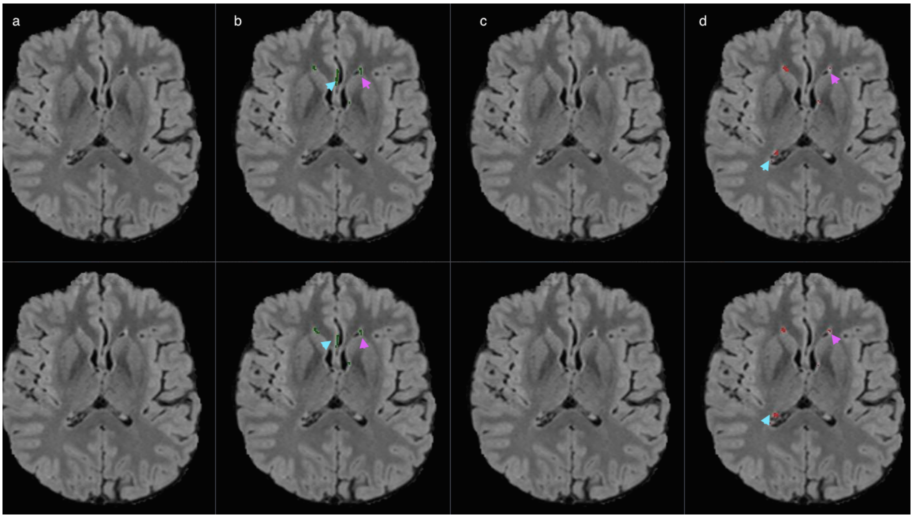


Figure A.5: Bias corrected FLAIR image (a) and super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) LST, (d) Lesion-TOADS. The first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2. Cyan arrow heads show false lesions detection for MS**metrix**-cross and Lesion-TOADS. Pink arrow heads show subtle lesions that are picked up by MS**metrix**-cross and Lesion-TOADS.

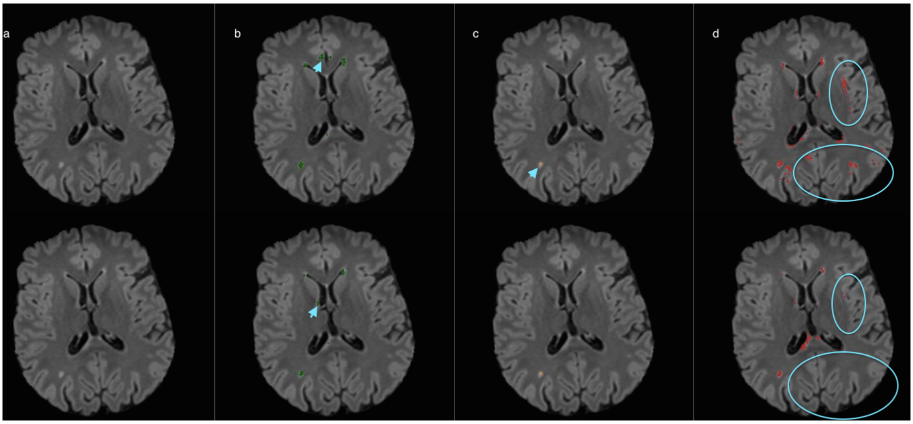


Figure A.6: Bias corrected FLAIR image (a) and super-imposed lesion segmentation from: (b) MS**metrix**-cross, (c) LST, (d) Lesion-TOADS. The first row corresponds to the lesion segmentation of scan 1 and the second row corresponds to the lesion segmentation of scan 2. Cyan ellipses represent the non-overlapping lesions between scan 1 and scan 2 for Lesion-TOADS, which are probably false positives. Cyan arrow heads show false lesion detection by MS**metrix**-cross and imprecise lesion boundary estimation by LST in both scans.

0.40). The lower Dice similarity index for Lesion-TOADS is primarily due to the fact that quite some non-overlapping lesions are found, which are probably false positives (cyan ellipse). For LST, the higher Dice similarity index is mainly because it consistently finds lesions across the scans. A low Dice similarity index for MS**metrix**-cross is mainly due to some false lesion detection in either of the scans (marked by cyan arrow heads). Although the Dice similarity index is higher for LST compared to MS**metrix**-cross, LST is slightly imprecise in lesion boundary estimation (marked by cyan arrow heads) for this case.

Bibliography

[1] Adalsteinsson, E., Irarrazabal, P., Topp, S., Meyer, C., Macovski, A., and Spielman, D. M. Volumetric spectroscopic imaging with spiral-based k-space trajectories. *Magnetic resonance in medicine 39*, 6 (1998), 889–898.

[2] Aït-Ali, L. S., Prima, S., Hellier, P., Carsin, B., Edan, G., and Barillot, C. STREM: a robust multidimensional parametric method to segment MS lesions in MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2005), Springer, pp. 409–416.

[3] Altay, E. E., Fisher, E., Jones, S. E., Hara-Cleaver, C., Lee, J.-C., and Rudick, R. A. Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic. *JAMA neurology 70*, 3 (2013), 338–344.

[4] Andronesi, O. C., Gagoski, B. A., and Sorensen, A. G. Neurologic 3D MR spectroscopic imaging with low-power adiabatic pulses and fast spiral acquisition. *Radiology 262*, 2 (2012), 647–661.

[5] Arnold, D. L., Li, D., Hohol, M., Chakraborty, S., Chankowsky, J., Alikhani, K., Duquette, P., Bhan, V., Montanera, W., Rabinovitch, H., et al. Evolving role of MRI in optimizing the treatment of multiple sclerosis: Canadian Consensus recommendations. *Multiple Sclerosis Journal–Experimental, Translational and Clinical 1* (2015), 2055217315589775.

[6] Bakshi, R., Benedict, R. H., Bermel, R. A., Caruthers, S. D., Puli, S. R., Tjoa, C. W., Fabiano, A. J., and Jacobs, L. T2 hypointensity in the deep gray matter of patients with multiple sclerosis: a quantitative magnetic resonance imaging study. *Archives of Neurology 59*, 1 (2002), 62–68.

[7] Banerjee, S., Ozturk-Isik, E., Nelson, S. J., and Majumdar, S. Elliptical magnetic resonance spectroscopic imaging with GRAPPA for imaging brain tumors at 3 T. *Magnetic resonance imaging 27*, 10 (2009), 1319–1325.

[8] Bao, Y., and Maudsley, A. A. Improved reconstruction for MR spectroscopic imaging. *IEEE transactions on medical imaging 26*, 5 (2007), 686–695.

[9] Barker, P. B., Breiter, S. N., Soher, B. J., Chatham, J. C., Forder, J. R., Samphilipo, M. A., Magee, C. A., and Anderson, J. H. Quantitative proton spectroscopy of canine brain: in vivo and in vitro correlations. *Magnetic Resonance in Medicine 32*, 2 (1994), 157–163.

[10] Barkhof, F., Calabresi, P. A., Miller, D. H., and Reingold, S. C. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. *Nature Reviews Neurology 5*, 5 (2009), 256–266.

[11] Barkhof, F., Filippi, M., Miller, D. H., Scheltens, P., Campi, A., Polman, C. H., Comi, G., Ader, H. J., Losseff, N., and Valk, J. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain 120*, 11 (1997), 2059–2069.

[12] Benedict, R. H., and Zivadinov, R. Risk factors for and management of cognitive dysfunction in multiple sclerosis. *Nature Reviews Neurology 7*, 6 (2011), 332–342.

[13] Bermel, R. A., and Bakshi, R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. *The Lancet Neurology 5*, 2 (2006), 158–170.

[14] Bernardis, E., Pohl, K. M., and Davatzikos, C. Extracting evolving pathologies via spectral clustering. In *International Conference on Information Processing in Medical Imaging* (2013), Springer, pp. 680–691.

[15] Bertholdo, D., Watcharakorn, A., and Castillo, M. Brain proton magnetic resonance spectroscopy: introduction and overview. *Neuroimaging Clinics of North America 23*, 3 (2013), 359–380.

[16] Blüml, S. Magnetic resonance spectroscopy: basics. In *MR Spectroscopy of Pediatric Brain Disorders*. Springer, 2013, pp. 11–23.

[17] Blystad, I., Håkansson, I., Tisell, A., Ernerudh, J., Smedby, Ö., Lundberg, P., and Larsson, E.-M. Quantitative MRI for Analysis of Active Multiple Sclerosis Lesions without Gadolinium-Based Contrast Agent. *American Journal of Neuroradiology 37*, 1 (2016), 94–100.

[18] Bogner, W., Gruber, S., Trattnig, S., and Chmelik, M. High-resolution mapping of human brain metabolites by free induction decay 1H MRSI at 7 T. *NMR in Biomedicine 25*, 6 (2012), 873–882.

[19] Bogner, W., Hess, A. T., Gagoski, B., Tisdall, M. D., van der Kouwe, A. J., Trattnig, S., Rosen, B., and Andronesi, O. C. Real-time motion-and B0-correction for LASER-localized spiral-accelerated 3D-MRSI of the brain at 3T. *NeuroImage 88* (2014), 22–31.

[20] Bosc, M., Heitz, F., Armspach, J.-P., Namer, I., Gounot, D., and Rumbach, L. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *NeuroImage 20*, 2 (2003), 643–656.

[21] Boyes, R. G., Rueckert, D., Aljabar, P., Whitwell, J., Schott, J. M., Hill, D. L., and Fox, N. C. Cerebral atrophy measurements using Jacobian integration: comparison with the boundary shift integral. *NeuroImage 32*, 1 (2006), 159–169.

[22] Brex, P. A., Ciccarelli, O., O'Riordan, J. I., Sailer, M., Thompson, A. J., and Miller, D. H. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New England Journal of Medicine 346*, 3 (2002), 158–164.

[23] Buades, A., Coll, B., and Morel, J.-M. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation 4*, 2 (2005), 490–530.

[24] Calabrese, M., Rinaldi, F., Grossi, P., and Gallo, P. Cortical pathology and cognitive impairment in multiple sclerosis. *Expert review of neurotherapeutics 11*, 3 (2011), 425–432.

[25] Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C. H., Cardoso, M. J., Cawley, N., Ciccarelli, O., Wheeler-Kingshott, C. A., Ourselin, S., Catanese, L., Deshpande, H., Maurel, P., Commowick, O., Barillot, C., Tomas-Fernandez, X., Warfield, S. K., Vaidya, S., Chunduru, A., Muthuganapathy, R., Krishnamurthi, G., Jesson, A., Arbel, T., Maier, O., Handels, H., Iheme, L. O., Unay, D., Jain, S., Sima, D. M., Smeets, D., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Bazin, P.-L., Calabresi, P. A., Crainiceanu, C. M., Ellingsen, L. M., Reich, D. S., Prince, J. L., and Pham, D. L. Longitudinal multiple sclerosis lesion segmentation: resource & challenge. *NeuroImage 148* (2017), 77–102.

[26] CARDOSO, M. Niftyseg-statistical segmentation and label fusion software package, 2012.

[27] CASTLEMAN, K. Digital image processing.

[28] CHARD, D. T., JACKSON, J. S., MILLER, D. H., AND WHEELER-KINGSHOTT, C. A. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *Journal of magnetic resonance imaging 32*, 1 (2010), 223–228.

[29] CIFELLI, A., ARRIDGE, M., JEZZARD, P., ESIRI, M. M., PALACE, J., AND MATTHEWS, P. M. Thalamic neurodegeneration in multiple sclerosis. *Annals of neurology 52*, 5 (2002), 650–653.

[30] COCOSCO, C. A., KOLLOKIAN, V., KWAN, R. K.-S., PIKE, G. B., AND EVANS, A. C. Brainweb: Online interface to a 3D MRI simulated brain database. In *NeuroImage* (1997), Citeseer.

[31] COHEN, J. Statistical power analysis for the behavior science. *Lawrance Eribaum Association* (1988).

[32] COOK, S. D. *Handbook of multiple sclerosis.* CRC Press, 2001.

[33] COTTON, F., WEINER, H. L., JOLESZ, F. A., AND GUTTMANN, C. R. MRI contrast uptake in new lesions in relapsing-remitting MS followed at weekly intervals. *Neurology 60*, 4 (2003), 640–646.

[34] COUPÉ, P., MANJÓN, J. V., CHAMBERLAND, M., DESCOTEAUX, M., AND HIBA, B. Collaborative patch-based super-resolution for diffusion-weighted images. *NeuroImage 83* (2013), 245–261.

[35] CRAMÉR, H. *Mathematical Methods of Statistics (PMS-9)*, vol. 9. Princeton university press, 2016.

[36] CUTTER, G. R., BAIER, M. L., RUDICK, R. A., COOKFAIR, D. L., FISCHER, J. S., PETKAU, J., SYNDULKO, K., WEINSHENKER, B. G., ANTEL, J. P., CONFAVREUX, C., ET AL. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain 122*, 5 (1999), 871–882.

[37] DAVIE, C., BARKER, G., THOMPSON, A., TOFTS, P., MCDONALD, W., AND MILLER, D. 1H magnetic resonance spectroscopy of chronic cerebral white matter lesions and normal appearing white matter in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry 63*, 6 (1997), 736–742.

[38] De Graaf, R. A. *In vivo NMR spectroscopy: principles and techniques.* John Wiley & Sons, 2013.

[39] De Stefano, N., Airas, L., Grigoriadis, N., Mattle, H. P., O'Riordan, J., Oreja-Guevara, C., Sellebjerg, F., Stankoff, B., Walczak, A., Wiendl, H., et al. Clinical relevance of brain volume measures in multiple sclerosis. *CNS drugs 28*, 2 (2014), 147–156.

[40] De Stefano, N., Giorgio, A., Battaglini, M., Rovaris, M., Sormani, M., Barkhof, F., Korteweg, T., Enzinger, C., Fazekas, F., Calabrese, M., et al. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. *Neurology 74*, 23 (2010), 1868–1876.

[41] Deeks, E. D. Dimethyl fumarate: a review in relapsing-remitting MS. *Drugs 76*, 2 (2016), 243–254.

[42] Despotović, I., Goossens, B., and Philips, W. MRI segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine 2015* (2015).

[43] Dice, L. R. Measures of the amount of ecologic association between species. *Ecology 26*, 3 (1945), 297–302.

[44] Dobson, R., Rudick, R. A., Turner, B., Schmierer, K., and Giovannoni, G. Assessing treatment response to interferon-β Is there a role for MRI? *Neurology 82*, 3 (2014), 248–254.

[45] Dreher, W., Erhard, P., and Leibfritz, D. Fast three-dimensional proton spectroscopic imaging of the human brain at 3T by combining spectroscopic missing pulse steady-state free precession and echo planar spectroscopic imaging. *Magnetic resonance in medicine 66*, 6 (2011), 1518–1525.

[46] Durand-Dubief, F., Belaroussi, B., Armspach, J., Dufour, M., Roggerone, S., Vukusic, S., Hannoun, S., Sappey-Marinier, D., Confavreux, C., and Cotton, F. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. *American Journal of Neuroradiology 33*, 10 (2012), 1918–1924.

[47] Dydak, U., Weiger, M., Pruessmann, K. P., Meier, D., and Boesiger, P. Sensitivity-encoded spectroscopic imaging. *Magnetic resonance in medicine 46*, 4 (2001), 713–722.

[48] Elliott, C., Arnold, D. L., Collins, D. L., and Arbel, T. Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI. *IEEE transactions on medical imaging 32*, 8 (2013), 1490–1503.

[49] Filippi, M., and Agosta, F. Magnetization transfer MRI in multiple sclerosis. *Journal of Neuroimaging 17*, s1 (2007), 22S–26S.

[50] Filippi, M., and Agosta, F. Imaging biomarkers in multiple sclerosis. *Journal of Magnetic Resonance Imaging 31*, 4 (2010), 770–788.

[51] Filippi, M., Horsfield, M., Tofts, P., Barkhof, F., Thompson, A., and Miller, D. Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain 118*, 6 (1995), 1601–1612.

[52] Filippi, M., Horsfield, M. A., Rovaris, M., Yousry, T., Rocca, M. A., Baratti, C., Bressi, S., and Comi, G. Intraobserver and interobserver variability in schemes for estimating volume of brain lesions on mr images in multiple sclerosis. *American journal of neuroradiology 19*, 2 (1998), 239–244.

[53] Filippi, M., Paty, D., Kappos, L., Barkhof, F., Compston, D., Thompson, A., Zhao, G., Wiles, C., McDonald, W., and Miller, D. Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis A follow-up study. *Neurology 45*, 2 (1995), 255–260.

[54] Fischl, B. FreeSurfer. *Neuroimage 62*, 2 (2012), 774–781.

[55] Fisher, E., Rudick, R., Cutter, G., Baier, M., Miller, D., Weinstock-Guttman, B., Mass, M., Dougherty, D., and Simonian, N. Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. *Multiple Sclerosis 6*, 6 (2000), 373–377.

[56] Fisher, E., Rudick, R., Simon, J., Cutter, G., Baier, M., Lee, J.-C., Miller, D., Weinstock-Guttman, B., Mass, M., Dougherty, D., et al. Eight-year follow-up study of brain atrophy in patients with MS. *Neurology 59*, 9 (2002), 1412–1420.

[57] Fisniku, L., Brex, P., Altmann, D., Miszkiel, K., Benton, C., Lanyon, R., Thompson, A., and Miller, D. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain 131*, 3 (2008), 808–817.

[58] Fox, R. J., and Cohen, J. A. Multiple sclerosis: the importance of early recognition and treatment. *Cleveland Clinic journal of medicine 68*, 2 (2001), 157–171.

[59] Freeborough, P. A., and Fox, N. C. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE transactions on medical imaging 16*, 5 (1997), 623–629.

[60] Freedman, M. S., Selchen, D., Arnold, D. L., Prat, A., Banwell, B., Yeung, M., Morgenthau, D., Lapierre, Y., Group, C. M. S. W., et al. Treatment optimization in MS: Canadian MS Working Group updated recommendations. *Canadian Journal of Neurological Sciences 40*, 3 (2013).

[61] García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical image analysis 17*, 1 (2013), 1–18.

[62] Ge, Y., Gonen, O., Inglese, M., Babb, J., Markowitz, C., and Grossman, R. Neuronal cell injury precedes brain atrophy in multiple sclerosis. *Neurology 62*, 4 (2004), 624–627.

[63] Geremia, E., Clatz, O., Menze, B. H., Konukoglu, E., Criminisi, A., and Ayache, N. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage 57*, 2 (2011), 378–390.

[64] Gerig, G., Welti, D., Guttmann, C. R., Colchester, A. C., and Székely, G. Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. *Medical Image Analysis 4*, 1 (2000), 31–42.

[65] Geurts, J. J., Calabrese, M., Fisher, E., and Rudick, R. A. Measurement and clinical effect of grey matter pathology in multiple sclerosis. *The Lancet Neurology 11*, 12 (2012), 1082–1092.

[66] Giorgio, A., and De Stefano, N. Cognition in multiple sclerosis: relevance of lesions, brain atrophy and proton MR spectroscopy. *Neurological sciences 31*, 2 (2010), 245–248.

[67] Giorgio, A., and De Stefano, N. Clinical use of brain volumetry. *Journal of Magnetic Resonance Imaging 37*, 1 (2013), 1–14.

[68] Giorgio, A., Stromillo, M. L., Bartolozzi, M. L., Rossi, F., Battaglini, M., De Leucio, A., Guidi, L., Maritato, P.,

PORTACCIO, E., SORMANI, M. P., ET AL. Relevance of hypointense brain MRI lesions for long-term worsening of clinical disability in relapsing multiple sclerosis. *Multiple Sclerosis Journal 20*, 2 (2014), 214–219.

[69] GIOVANNONI, G., TURNER, B., GNANAPAVAN, S., OFFIAH, C., SCHMIERER, K., AND MARTA, M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? *Multiple sclerosis and related disorders 4*, 4 (2015), 329–333.

[70] GOVINDARAJU, V., YOUNG, K., AND MAUDSLEY, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine 13*, 3 (2000), 129–153.

[71] GRISWOLD, M. A., JAKOB, P. M., HEIDEMANN, R. M., NITTKA, M., JELLUS, V., WANG, J., KIEFER, B., AND HAASE, A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic resonance in medicine 47*, 6 (2002), 1202–1210.

[72] HALDAR, J. P., HERNANDO, D., SONG, S.-K., AND LIANG, Z.-P. Anatomically constrained reconstruction from noisy data. *Magnetic Resonance in Medicine 59*, 4 (2008), 810–818.

[73] HANGEL, G., STRASSER, B., POVAŽAN, M., GRUBER, S., CHMELÍK, M., GAJDOŠÍK, M., TRATTNIG, S., AND BOGNER, W. Lipid suppression via double inversion recovery with symmetric frequency sweep for robust 2D-GRAPPA-accelerated MRSI of the brain at 7 T. *NMR in Biomedicine 28*, 11 (2015), 1413–1425.

[74] HANGEL, G., STRASSER, B., POVAŽAN, M., HECKOVA, E., HINGERL, L., BOUBELA, R., GRUBER, S., TRATTNIG, S., AND BOGNER, W. Ultra-high resolution brain metabolite mapping at 7 T by short-TR Hadamard-encoded FID-MRSI. *NeuroImage* (2016), in press.

[75] HOWARD, J., TREVICK, S., AND YOUNGER, D. S. Epidemiology of Multiple Sclerosis. *Neurologic Clinics 34*, 4 (2016), 919–939.

[76] HU, X., LEVIN, D. N., LAUTERBUR, P. C., AND SPRAGGINS, T. SLIM: Spectral localization by imaging. *Magnetic resonance in medicine 8*, 3 (1988), 314–322.

[77] HYLAND, M., AND RUDICK, R. A. Challenges to clinical trials in multiple sclerosis: outcome measures in the era of disease-modifying drugs. *Current opinion in neurology 24*, 3 (2011), 255–261.

[78] IGLESIAS, J. E., LIU, C.-Y., THOMPSON, P. M., AND TU, Z. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE transactions on medical imaging 30*, 9 (2011), 1617–1634.

[79] Inglese, M., Li, B. S., Rusinek, H., Babb, J. S., Grossman, R. I., and Gonen, O. Diffusely elevated cerebral choline and creatine in relapsing-remitting multiple sclerosis. *Magnetic resonance in medicine 50*, 1 (2003), 190–195.

[80] Jacob, M., Zhu, X., Ebel, A., Schuff, N., and Liang, Z.-P. Improved model-based magnetic resonance spectroscopic imaging. *IEEE transactions on medical imaging 26*, 10 (2007), 1305–1318.

[81] Jacobsen, C., Hagemeier, J., Myhr, K.-M., Nyland, H., Lode, K., Bergsland, N., Ramasamy, D. P., Dalaker, T. O., Larsen, J. P., Farbu, E., et al. Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. *Journal of Neurology, Neurosurgery & Psychiatry* (2014), jnnp–2013.

[82] Jäger, F., Deuerling-Zheng, Y., Frericks, B., Wacker, F., and Hornegger, J. A new method for MRI intensity standardization with application to lesion detection in the brain. *Procs 1010* (2006), 269–276.

[83] Jain, S., Ribbens, A., Sima, D. M., Cambron, M., De Keyser, J., Wang, C., Barnett, M. H., Van Huffel, S., Maes, F., and Smeets, D. Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework. *Frontiers in Neuroscience:Brain Imaging Methods 10* (2016), 576.

[84] Jain, S., Sima, D., Sanaei Nezhad, F., Hangel, G., Bogner, W., Williams, S., Van Huffel, S., Maes, F., and Smeets, D. Patch-based super-resolution of MR spectroscopic images: Application to Multiple Sclerosis. *Frontiers in Neuroscience:Brain Imaging Methods 11* (2017), 13.

[85] Jain, S., Sima, D. M., Nezhad, F. S., Williams, S., Huffel, S. V., Maes, F., and Smeets, D. Patch based super-resolution of MR spectroscopic images. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (April 2016), pp. 452–456.

[86] Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., Daams, M., et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical 8* (2015), 367–375.

[87] Jansen, J. F., Backes, W. H., Nicolay, K., and Kooi, M. E. 1h mr spectroscopy of the brain: absolute quantification of metabolites. *Radiology 240*, 2 (2006), 318–332.

[88] Kappos, L., De Stefano, N., Freedman, M. S., Cree, B. A., Radue, E.-W., Sprenger, T., Sormani, M. P., Smith, T., Häring, D. A., Meier, D. P., et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity ' (neda-4) in relapsing–remitting multiple sclerosis. *Multiple Sclerosis Journal* (2015), 1352458515616701.

[89] Kasten, J., Klauser, A., Lazeyras, F., and Van De Ville, D. Magnetic resonance spectroscopic imaging at superresolution: Overview and perspectives. *Journal of Magnetic Resonance 263* (2016), 193–208.

[90] Khalidov, I., Van De Ville, D., Jacob, M., Lazeyras, F., and Unser, M. BSLIM: Spectral localization by imaging with explicit field inhomogeneity compensation. *IEEE transactions on medical imaging 26*, 7 (2007), 990–1000.

[91] Khayati, R., Vafadust, M., Towhidkhah, F., and Nabavi, M. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. *Computers in biology and medicine 38*, 3 (2008), 379–390.

[92] Kirchner, T., Fillmer, A., Tsao, J., Pruessmann, K. P., and Henning, A. Reduction of voxel bleeding in highly accelerated parallel 1H MRSI by direct control of the spatial response function. *Magnetic resonance in medicine 73*, 2 (2015), 469–480.

[93] Koch, G. G. Intraclass correlation coefficient. *Encyclopedia of statistical sciences* (1982).

[94] Kornak, J., Young, K., Soher, B. J., and Maudsley, A. A. Bayesian-Space–Time Reconstruction of MR Spectroscopic Imaging for Enhanced Resolution. *IEEE transactions on medical imaging 29*, 7 (2010), 1333–1350.

[95] Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology 33*, 11 (1983), 1444–1444.

[96] Lao, Z., Shen, D., Liu, D., Jawad, A. F., Melhem, E. R., Launer, L. J., Bryan, R. N., and Davatzikos, C. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Academic radiology 15*, 3 (2008), 300–313.

[97] Lavery, A. M., Verhey, L. H., and Waldman, A. T. Outcome measures in relapsing-remitting multiple sclerosis: capturing disability and disease progression in clinical trials. *Multiple sclerosis international 2014:262350* (2014).

[98] Lewis, E. B., and Fox, N. C. Correction of differential intensity inhomogeneity in longitudinal MR images. *Neuroimage 23*, 1 (2004), 75–83.

[99] Liang, Z.-P., and Lauterbur, P. C. A generalized series approach to MR spectroscopic imaging. *IEEE transactions on medical imaging 10*, 2 (1991), 132–137.

[100] Lin, F.-H., Tsai, S.-Y., Otazo, R., Caprihan, A., Wald, L. L., Belliveau, J. W., and Posse, S. Sensitivity-encoded (SENSE) proton echo-planar spectroscopic imaging (PEPSI) in the human brain. *Magnetic resonance in medicine 57*, 2 (2007), 249–257.

[101] Lladó, X., Ganiler, O., Oliver, A., Martí, R., Freixenet, J., Valls, L., Vilanova, J. C., Ramió-Torrentà, L., and Rovira, À. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology 54*, 8 (2012), 787–807.

[102] Ma, C., Lam, F., Ning, Q., Johnson, C. L., and Liang, Z.-P. High-resolution 1H-MRSI of the brain using short-TE SPICE. *Magnetic resonance in medicine* (2016).

[103] Manjón, J. V., Coupé, P., Buades, A., Collins, D. L., and Robles, M. MRI superresolution using self-similarity and image priors. *Journal of Biomedical Imaging 2010* (2010), 17.

[104] Manjón, J. V., Coupé, P., Buades, A., Fonov, V., Collins, D. L., and Robles, M. Non-local MRI upsampling. *Medical image analysis 14*, 6 (2010), 784–792.

[105] Miller, D. H. Biomarkers and surrogate outcomes in neurodegenerative disease: lessons from multiple sclerosis. *NeuroRx 1*, 2 (2004), 284–294.

[106] Mitjana, R., Tintoré, M., Rocca, M. A., Auger, C., Barkhof, F., Filippi, M., Polman, C., Fazekas, F., Huerga, E., Montalban, X., et al. Diagnostic value of brain chronic black holes on T1-weighted MR images in clinically isolated syndromes. *Multiple Sclerosis Journal 20*, 11 (2014), 1471–1477.

[107] Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., and Ourselin, S. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine 98*, 3 (2010), 278–284.

[108] Mortazavi, D., Kouzani, A. Z., and Soltanian-Zadeh, H. Segmentation of multiple sclerosis lesions in MR images: a review. *Neuroradiology 54*, 4 (2012), 299–320.

[109] Multiple Sclerosis International Federation Atlas of MS. www.atlasofms.org, online; accessed 30 May 2017.

[110] Nagtegaal, G. J., Pohl, C., Wattjes, M. P., Hulst, H. E., Freedman, M. S., Hartung, H.-P., Miller, D., Montalban, X., Kappos, L., Edan, G., et al. Interferon beta-1b reduces black holes in a randomised trial of clinically isolated syndrome. *Multiple Sclerosis Journal 20*, 2 (2014), 234–242.

[111] Narayana, P. A., Doyle, T. J., Lai, D., and Wolinsky, J. S. Serial proton magnetic resonance spectroscopic imaging, contrast-enhanced magnetic resonance imaging, and quantitative lesion volumetry in multiple sclerosis. *Annals of neurology 43*, 1 (1998), 56–71.

[112] Naressi, A., Couturier, C., Castang, I., De Beer, R., and Graveron-Demilly, D. Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. *Computers in biology and medicine 31*, 4 (2001), 269–286.

[113] National Multiple Sclerosis Society. http://www.nationalmssociety.org/What-is-MS/Types-of-MS, online; accessed 02 Feburary 2017.

[114] O'riordan, J., Thompson, A., Kingsley, D., MacManus, D., Kendall, B., Rudge, P., McDonald, W., and Miller, D. The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain 121*, 3 (1998), 495–503.

[115] Otazo, R., Tsai, S.-Y., Lin, F.-H., and Posse, S. Accelerated short-TE 3D proton echo-planar spectroscopic imaging using 2D-SENSE with a 32-channel array coil. *Magnetic resonance in medicine 58*, 6 (2007), 1107–1116.

[116] Ourselin, S., Stefanescu, R., and Pennec, X. Robust registration of multi-modal images: towards real-time clinical applications. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2002), Springer, pp. 140–147.

[117] Petzold, A., Wattjes, M. P., Costello, F., Flores-Rivera, J., Fraser, C. L., Fujihara, K., Leavitt, J., Marignier, R., Paul, F., Schippling, S., et al. The investigation of acute optic neuritis: a review and proposed protocol. *Nature Reviews Neurology 10*, 8 (2014), 447–458.

[118] Pohmann, R., Von Kienlin, M., and Haase, A. Theoretical evaluation and comparison of fast chemical shift imaging methods. *Journal of Magnetic Resonance 129*, 2 (1997), 145–160.

[119] Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology 69*, 2 (2011), 292–302.

[120] Popescu, V., Agosta, F., Hulst, H. E., Sluimer, I. C., Knol, D. L., Sormani, M. P., Enzinger, C., Ropele, S., Alonso, J., Sastre-Garriga, J., et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry 84*, 10 (2013), 1082–1091.

[121] Posse, S., Otazo, R., Dager, S. R., and Alger, J. MR spectroscopic imaging: principles and recent advances. *Journal of Magnetic Resonance Imaging 37*, 6 (2013), 1301–1325.

[122] Posse, S., Tedeschi, G., Risinger, R., Ogg, R., and Bihan, D. L. High Speed 1H Spectroscopic Imaging in Human Brain by Echo Planar Spatial-Spectral Encoding. *Magnetic resonance in medicine 33*, 1 (1995), 34–40.

[123] Považan, M., Hangel, G., Strasser, B., Gruber, S., Chmelik, M., Trattnig, S., and Bogner, W. Mapping of brain macromolecules and their use for spectral processing of 1 H-MRSI data with an ultra-short acquisition delay at 7T. *NeuroImage 121* (2015), 126–135.

[124] Prima, S., Ayache, N., Janke, A., Francis, S. J., Arnold, D. L., and Collins, D. L. Statistical analysis of longitudinal MRI data: applications for detection of disease activity in MS. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2002), Springer, pp. 363–371.

[125] Prosperini, L., Gallo, V., Petsas, N., Borriello, G., and Pozzilli, C. One-year MRI scan predicts clinical response to interferon beta in multiple sclerosis. *European journal of neurology 16*, 11 (2009), 1202–1209.

[126] Provencher, S. W. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic resonance in medicine 30*, 6 (1993), 672–679.

[127] Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. SENSE: sensitivity encoding for fast MRI. *Magnetic resonance in medicine 42*, 5 (1999), 952–962.

[128] Quarantelli, M. MRI/MRS in neuroinflammation: methodology and applications. *Clinical and translational imaging 3*, 6 (2015), 475–489.

[129] Rao, S., Martin, A., Huelin, R., Wissinger, E., Khankhel, Z., Kim, E., and Fahrbach, K. Correlations between MRI and information processing speed in MS: A meta-analysis. *Multiple sclerosis international 2014:975803* (2014).

[130] Rey, D., Subsol, G., Delingette, H., and Ayache, N. Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. *Medical image analysis 6*, 2 (2002), 163–179.

[131] Richards, J. E., Sanchez, C., Phillips-Meek, M., and Xie, W. A database of age-appropriate average MRI templates. *Neuroimage 124* (2016), 1254–1259.

[132] Rio, J., Castillo, J., Rovira, A., Tintore, M., Sastre-Garriga, J., Horga, A., Nos, C., Comabella, M., Aymerich, X., and Montalbán, X. Measures in the first year of therapy predict the response to interferon β in ms. *Multiple Sclerosis Journal 15*, 7 (2009), 848–853.

[133] Rio, J., Rovira, A., Tintoré, M., Huerga, E., Nos, C., Tellez, N., Tur, C., Comabella, M., and Montalban, X. Relationship between MRI lesion activity and response to IFN-β in relapsing–remitting multiple sclerosis patients. *Multiple Sclerosis 14*, 4 (2008), 479–484.

[134] Roussas, G. G. Nonparametric regression estimation under mixing conditions. *Stochastic processes and their applications 36*, 1 (1990), 107–116.

[135] Rousseau, F., Initiative, A. D. N., et al. A non-local approach for image super-resolution using intermodality priors. *Medical image analysis 14*, 4 (2010), 594–605.

[136] Rovira, À., Auger, C., and Alonso, J. Magnetic resonance monitoring of lesion evolution in multiple sclerosis. *Therapeutic advances in neurological disorders 6*, 5 (2013), 298–310.

[137] Rovira, À., Wattjes, M. P., Tintoré, M., Tur, C., Yousry, T. A., Sormani, M. P., De Stefano, N., Filippi, M., Auger, C.,

Rocca, M. A., et al. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis — clinical implementation in the diagnostic process. *Nature Reviews Neurology* (2015).

[138] Roxburgh, R., Seaman, S., Masterman, T., Hensiek, A., Sawcer, S., Vukusic, S., Achiti, I., Confavreux, C., Coustans, M., Le Page, E., et al. Multiple Sclerosis Severity Score Using disability and disease duration to rate disease severity. *Neurology 64*, 7 (2005), 1144–1151.

[139] Sajja, B. R., Wolinsky, J. S., and Narayana, P. A. Proton magnetic resonance spectroscopy in multiple sclerosis. *Neuroimaging clinics of North America 19*, 1 (2009), 45–58.

[140] Sarchielli, P., Presciutti, O., Tarducci, R., Gobbi, G., Alberti, A., Pelliccioli, G., Chiarini, P., and Gallai, V. Localized 1h magnetic resonance spectroscopy in mainly cortical gray matter of patients with multiple sclerosis. *Journal of neurology 249*, 7 (2002), 902–910.

[141] Sastre Garriga, J., et al. Brain volume and brain metabolite changes in the first stages of primary progressive multiple sclerosis.

[142] Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage 59*, 4 (2012), 3774–3783.

[143] Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J., and Miller, D. H. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Annals of neurology 56*, 3 (2004), 407–415.

[144] Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., and Leahy, R. M. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage 13*, 5 (2001), 856–876.

[145] Shi, J., Baxter, L. C., and Kuniyoshi, S. M. Pathologic and Imaging Correlates of Cognitive Deficits in Multiple Sclerosis: Changing the Paradigm of Diagnosis and Prognosis. *Cognitive and Behavioral Neurology 27*, 1 (2014), 1–7.

[146] Shiee, N., Bazin, P.-L., Ozturk, A., Reich, D. S., Calabresi, P. A., and Pham, D. L. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage 49*, 2 (2010), 1524–1535.

[147] Shiee, N., Bazin, P.-L., Zackowski, K. M., Farrell, S. K., Harrison, D. M., Newsome, S. D., Ratchford, J. N., Caffo, B. S., Calabresi, P. A., Pham, D. L., et al. Revisiting brain atrophy and its relationship to disability in multiple sclerosis. *PLoS One 7*, 5 (2012), e37049.

[148] Sicotte, N. L. Magnetic resonance imaging in multiple sclerosis: the role of conventional imaging. *Neurologic clinics 29*, 2 (2011), 343–356.

[149] Sima, D. M., Loeckx, D., Smeets, D., Jain, S., Parizel, P. M., and Van Hecke, W. Use Case I: Imaging Biomarkers in Neurological Disease. Focus on Multiple Sclerosis. In *Imaging Biomarkers*. Springer, 2017, pp. 169–180.

[150] Sled, J. G., Zijdenbos, A. P., and Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging 17*, 1 (1998), 87–97.

[151] Smeets, D., Ribbens, A., Sima, D. M., Cambron, M., Horakova, D., Jain, S., Maertens, A., Van Vlierberghe, E., Terzopoulos, V., Van Binst, A.-M., et al. Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain and behavior 6*, 9 (2016).

[152] Smith, S. M. Fast robust automated brain extraction. *Human brain mapping 17*, 3 (2002), 143–155.

[153] Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., and De Stefano, N. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage 17*, 1 (2002), 479–489.

[154] Solomon, J., and Sood, A. 4-D lesion detection using expectation-maximization and hidden Markov model. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on* (2004), pp. 125–128.

[155] Sormani, M., Bonzano, L., Roccatagliata, L., Mancardi, G., Uccelli, A., and Bruzzi, P. Surrogate endpoints for EDSS worsening in multiple sclerosis A meta-analytic approach. *Neurology 75*, 4 (2010), 302–309.

[156] Sormani, M. P., Arnold, D. L., and De Stefano, N. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Annals of neurology 75*, 1 (2014), 43–49.

[157] Sormani, M. P., Bonzano, L., Roccatagliata, L., Cutter, G. R., Mancardi, G. L., and Bruzzi, P. Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: A meta-analytic approach. *Annals of neurology 65*, 3 (2009), 268–275.

[158] Sormani, M. P., and Bruzzi, P. MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. *The Lancet Neurology 12*, 7 (2013), 669–676.

[159] Stangel, M., Penner, I. K., Kallmann, B. A., Lukas, C., and Kieseier, B. C. Towards the implementation of 'no evidence of disease activity' in multiple sclerosis treatment: the multiple sclerosis decision model. *Therapeutic advances in neurological disorders 8*, 1 (2015), 3–13.

[160] Steenwijk, M. D., Pouwels, P. J., Daams, M., van Dalen, J. W., Caan, M. W., Richard, E., Barkhof, F., and Vrenken, H. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *NeuroImage: Clinical 3* (2013), 462–469.

[161] Strasser, B., Chmelik, M., Robinson, S., Hangel, G., Gruber, S., Trattnig, S., and Bogner, W. Coil combination of multichannel MRSI data at 7 T: MUSICAL. *NMR in Biomedicine 26*, 12 (2013), 1796–1805.

[162] Strasser, B., Považan, M., Hangel, G., Hingerl, L., Chmelik, M., Gruber, S., Trattnig, S., and Bogner, W. (2+ 1) D-CAIPIRINHA accelerated MR spectroscopic imaging of the brain at 7T. *Magnetic Resonance in Medicine* (2016), in press.

[163] study group, M., et al. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis — establishing disease prognosis and monitoring patients. *Nature Reviews Neurology 11*, 10 (2015), 597–606.

[164] Suhy, J., Rooney, W., Goodkin, D., Capizzano, A., Soher, B., Maudsley, A. A., Waubant, E., Andersson, P., and Weiner, M. 1H MRSI comparison of white matter and lesions in primary progressive and relapsing-remitting MS. *Multiple sclerosis 6*, 3 (2000), 148–155.

[165] Székely, G., Welti, D., Gerig, G., Radü, E.-W., and Kappos, L. Spatio-temporal segmentation of active multiple sclerosis lesions in serial MRI data. In *Biennial International Conference on Information Processing in Medical Imaging* (2001), Springer, pp. 438–445.

[166] Tartaglia, M., Narayanan, S., De Stefano, N., Arnaoutelis, R., Antel, S., Francis, S., Santos, A., Lapierre, Y., and Arnold, D. Choline is increased in pre-lesional normal appearing white matter in multiple sclerosis. *Journal of neurology 249*, 10 (2002), 1382–1390.

[167] Tiberio, M., Chard, D., Altmann, D., Davies, G., Griffin, C., McLean, M., Rashid, W., Sastre-Garriga, J., Thompson, A., and Miller, D. Metabolite changes in early relapsing–remitting multiple sclerosis. *Journal of neurology 253*, 2 (2006), 224–230.

[168] Tintore, M., Rovira, A., Arrambide, G., Mitjana, R., Rio, J., Auger, C., Nos, C., Edo, M., Castillo, J., Horga, A., et al. Brainstem lesions in clinically isolated syndromes. *Neurology 75*, 21 (2010), 1933–1938.

[169] Traboulsee, A., Simon, J., Stone, L., Fisher, E., Jones, D., Malhotra, A., Newsome, S., Oh, J., Reich, D., Richert, N., et al. Revised recommendations of the consortium of ms centers task force for a standardized MRI protocol and clinical guidelines for the diagnosis and follow-up of multiple sclerosis. *American Journal of Neuroradiology 37*, 3 (2016), 394–401.

[170] Tur, C., Montalban, X., Tintoré, M., Nos, C., Río, J., Aymerich, F. X., Brieva, L., Téllez, N., Perkal, H., Comabella, M., et al. Interferon beta-1b for the treatment of primary progressive multiple sclerosis: five-year clinical trial follow-up. *Archives of neurology 68*, 11 (2011), 1421–1427.

[171] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging 29*, 6 (2010), 1310–1320.

[172] Vaina, L., and Jolesz, F. 4D connected component labelling applied to quantitative analysis of MS lesion temporal development.

[173] Van Cauter, S., Sima, D. M., Luts, J., ter Beek, L., Ribbens, A., Peeters, R. R., Osorio Garcia, M. I., Li, Y., Sunaert, S., Van Gool, S. W., et al. Reproducibility of rapid short echo time CSI at 3 Tesla for clinical applications. *Journal of Magnetic Resonance Imaging 37*, 2 (2013), 445–456.

[174] van der Graaf, M. In vivo magnetic resonance spectroscopy: basic methodology and clinical applications. *European Biophysics Journal 39*, 4 (2010), 527–540.

[175] van Der Graaf, M., Heerschap, A., et al. Common processing of in vivo MR spectra. *NMR in biomedicine 14*, 4 (2001), 224–232.

[176] Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., and Suetens, P. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE transactions on medical imaging 20*, 8 (2001), 677–688.

[177] Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. Automated model-based tissue classification of MR images of the brain. *IEEE transactions on medical imaging 18*, 10 (1999), 897–908.

[178] Van Waesberghe, J., Kamphorst, W., De Groot, C. J., Van Walderveen, M. A., Castelijns, J. A., Ravid, R., Lycklama a Nijeholt, G., Van Der Valk, P., Polman, C. H., Thompson, A. J., et al. Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability. *Annals of neurology 46*, 5 (1999), 747–754.

[179] Vrenken, H., Jenkinson, M., Horsfield, M., Battaglini, M., Van Schijndel, R., Rostrup, E., Geurts, J., Fisher, E., Zijdenbos, A., Ashburner, J., et al. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. *Journal of neurology 260*, 10 (2013), 2458–2471.

[180] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing 13*, 4 (2004), 600–612.

[181] Wattjes, M., Steenwijk, M., and Stangel, M. MRI in the diagnosis and monitoring of multiple sclerosis: an update. *Clinical neuroradiology 25*, 2 (2015), 157–165.

[182] Welch, B. L. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika 34*, 1/2 (1947), 28–35.

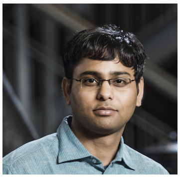
[183] Werring, D., Clark, C., Barker, G., Thompson, A., and Miller, D. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology 52*, 8 (1999), 1626–1626.

[184] Wolkorte, R., Heersema, D. J., and Zijdewind, I. Muscle fatigability during a sustained index finger abduction and depression scores are associated with perceived fatigue in patients with relapsing-remitting multiple sclerosis. *Neurorehabilitation and neural repair 29*, 8 (2015), 796–802.

[185] Zhang, Y., Brady, M., and Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging 20*, 1 (2001), 45–57.

[186] Zhu, H., Soher, B. J., Ouwerkerk, R., Schär, M., and Barker, P. B. Spin-echo magnetic resonance spectroscopic imaging at 7 T with frequency-modulated refocusing pulses. *Magnetic resonance in medicine 69*, 5 (2013), 1217–1225.

Curriculum vitae



Saurabh Jain (Ajmer, India, 2 September 1986) obtained his high school diploma at Saint Anselm's Senior Secondary School Ajmer, in 2004. In 2009, he graduated at the Indian Institute of Information technology, Design and Manufacturing, Jabalpur, India as Bachelor of Technology in Electronics and Communication Engineering. Till 2011, he was working as a developer in the R&D department of Hella Electronics Development Centre, Pune, India. In 2013, he completed his master in Computer Vision (Erasmus Mundus program Vision in Robotics (VIBOT)) from University of Dijon (France), University of Girona (Spain) and Heriot-Watt University (UK). The subject of his master thesis was 'Early diagnosis of Alzheimer disease', which he did at ico**metrix**, Leuven, Belgium in collaboration with KU Leuven under the supervision of Prof. dr. Dirk Vandermeulen (supervisor), Prof. dr. Frederik Maes, Dr. Dirk Loeckx (co-supervisors), Dr. Dirk Smeets and Dr. Annemie Ribbens (daily supervisors).

Since September 2013, he is working as an early stage researcher at ico**metrix** under the Marie Curie Initial Training Network program (TRANSACT) where he is making his PhD on *Combining anatomical and spectral information to enhance MRSI resolution and quantification* in collaboration with KU Leuven under the supervision of Prof. dr. Frederik Maes, Prof. dr. Sabine Van Huffel and Dr. Dirk Smeets. He co-authored 9 publications in international journals and international reviewed conference proceedings, one patent, and one book chapter. He also participated in MS lesion challenge 2015, where MS**metrix**-cross method was the best untrained method and was the among the top 5 algorithms. Additionally, he is also an active team member in the dynamic R&D team of ico**metrix** and is responsible for research and development of MS lesion segmentation software: MS**metrix**.

In his free time, Saurabh enjoys travelling, watching movies, listening music and meditation.

List of Publications

International Journal: Published

1. **Jain, S.**, Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., Daams, M., Maes, F., Van Huffel, S., Vrenken, H., and Smeets, D., Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical 8* (2015), 367–375.

2. **Jain, S.**, Ribbens, A., Sima, D.M., Cambron, M., De Keyser, J., Wang, C., Barnett, M.H., Van Huffel, S., Maes, F. and Smeets, D., Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework. *Frontiers in Neuroscience 10* (2016), 576.

3. **Jain, S.**, Sima, D.M., Sanaei Nezhad, F., Hangel, G., Bogner, W., Williams, S., Van Huffel, S., Maes, F. and Smeets, D., Patch-based super-resolution of MR spectroscopic images: Application to Multiple Sclerosis *Frontiers in Neuroscience 11* (2017), 13.

4. Smeets, D., Ribbens, A., Sima, D. M., Cambron, M., Horakova, D., **Jain, S.**, Maertens, A., Van Vlierberghe, E., Terzopoulos, V., Van Binst, A.M., Vaneckova M., Krasensky, J., Uher, T., Seidl, Z., De Keyser, J., Nagels, G., De Mey, J., Havrdova, E., and Van Hecke, W., Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain and behavior 6, 9* (2016).

5. Carass, A., Roy, S., Jog, A., Cuzzocreo, J.L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C.H., Cardoso, M.J., Cawley, N., Ciccarelli, O., Wheeler-Kingshott, C.A.M., Ourselin, S.,Catanesei, L., Deshpande, H., Maureli, P., Commowick, O., Barillot, C., Tomas-Fernandez, X., Warfield, S.K., Vaidya, S., Chunduru, A., Muthuganapathy, R., Krishnamurthi, G., Jesson, A., Arbel, T., Maier, O., Handels H.,

Iheme, L.O., Unay, D., **Jain, S.**, Sima, D.M., Smeets, D., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Bazin, P-L., Calabresi,P.A., Crainiceanu, C.M., Ellingsena, L.M., Reichd, D.S., Prince, J.L., and Pham D.L., Longitudinal multiple sclerosis lesion segmentation: resource & challenge, *NeuroImage 148* (2017), 77–102.

6. De Winter, F.L., Emsell, L., Bouckaert, F., Claes, L., **Jain, S.**, Farrar, G., Billiet, T., Evers, S., Van den Stock, J., Sienaert, P., Obbels, J., Sunaert, S., Adamczuk, K., Vandenberghe, R., Van Laere, K., and Vandenbulcke, M., No Association of Lower Hippocampal Volume With Alzheimer's Disease Pathology in Late-Life Depression. *American Journal of Psychiatry* (2016), appi-ajp.

7. Van Schependom, J., **Jain, S.**, Cambron, M., Van Binst, A.M., De Mey, J., Smeets, D., and Nagels, G., Reliability of measuring regional callosal atrophy in neurodegenerative diseases. *NeuroImage: Clinical, 12* (2016), 825–831.

Book chapter

Sima, D.M., Loeckx, D., Smeets, D., **Jain, S.**, Parizel, P.M., and Van Hecke, W., Use Case I: Imaging Biomarkers in Neurological Disease. Focus on Multiple Sclerosis, In *Imaging Biomarkers: Development and Clinical Integration*, Eds. Martí-Bonmatí, L. and Alberich-Bayarri, A., Springer International Publishing, 2016, 169-180.

International Conference or Workshop: Published in Proceedings

1. **Jain, S.**, Sima, D.M., Nezhad, F. S., Williams, S., Huffel, S. V., Maes, F., and Smeets, D. Patch based super-resolution of MR spectroscopic images, In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* (April 2016), 452–456.

2. **Jain, S.**, Ribbens A., Sima, D.M., Maes, F., and Smeets, D. Unsupervised framework for consistent longitudinal MS lesion segmentation, In *Bayesian and grAphical Models for Biomedical Imaging (BAMBI) workshop, Medical Image Computing and Computer Assisted Intervention (MICCAI)* (October 2016) (accepted).

International Conference: Abstract or Not Published

1. **Jain, S.**, Smeets, D., and Loeckx, D., Prior built population atlas for automatic hippocampus segmentation, ESMRMB 2013, abstract and electronic poster presentation, status-accepted.
2. **Jain, S.**, Smeets, D., Sima, D.M., Van Hecke, W., Loeckx, D., Van Huffel, S., and Maes, F., Multiple sclerosis brain lesion measurements in clinical practice, EFNS-ENS 2014, Joint Congress of European Neurology, abstract submission, status-accepted.
3. **Jain, S.**, Smeets, D., Ribbens, A., Sima, D.M., Janssens, K., Daams, M., Steenwijk, M., Vrenken, H., Barkhof, F., and Van Hecke, W., Automatic multiple sclerosis brain lesion localization and volumetry, ECTRIMS 2014, abstract and paper poster presentation, status-online.
4. **Jain, S.**, Smeets, D., Ribbens, A., Sima, D.M., Janssens, K., Daams, M., Steenwijk, M., Vrenken, H., Barkhof, F., and Van Hecke, W., Brain atrophy quantification in multiple sclerosis, ECTRIMS 2014, abstract and paper poster presentation, status-online.
5. **Jain, S.**, Sima, D.M., and Smeets, D., Automatic longitudinal Multiple Sclerosis lesion segmentation: MSmetrix, ISBI Longitudinal Multiple Sclerosis Lesion Segmentation Challenge, 2015, status-online.
6. **Jain, S.**, Sima, D.M., Maertens, A., Van Huffel, S., Maes, F., and Smeets, D., MSmetrix: accurate untrained method for longitudinal lesion segmentation, ECTRIMS 2015, abstract and paper poster presentation, status-online.
7. **Jain, S.**, Ribbens, A., Sima, D.M., Cambron, M., De Keyser, J., Wang, C., Barnett, M., Van Huffel, S., Maes, F., Maertens A., and Smeets D., Longitudinal assessment of brain lesion changes in follow-up MRI scans of patients with Multiple Sclerosis, ECTRIMS 2016, paper poster presentation, status-online
8. **Jain, S.**, Sima, D.M., Van Vlierberghe, E., Dubois, B., Dupont, P., Kocevar, G., Durand-Dubief, F., Sappey-Marinier, D., Wang, C., Barnett, M.H., Van Huffel, S., Maes, F. and Smeets, D. Clinical relevance of new and enlarging lesion volume in Relapsing Remitting Multiple Sclerosis: A multi-center study, The consortium of Multiple Sclerosis Centres 2017, abstract and paper poster presentation, status-accepted.

9. **Jain, S.**, Hangel, G., Sima, D.M., Bogner, W., Trattnig, S., Van Huffel, S., Maes, F. and Smeets, D. Patch-based super-resolution of MRSI data in multiple sclerosis patients at 7T, International Society for Magnetic Resonance in Medicine (ISMRM) 2017, abstract and paper poster presentation, status-accepted.
10. Ribbens, A., Cambron, M., De Keyser, J., Van Binst, A., De Mey, J., **Jain, S.**, Nagels G., Maertens A., and Van Hecke W., Quantifying brain atrophy in clinical practice for MS patients: a feasibility study on the measurement error, ECR 2015, status-online.
11. Ribbens, A., Smeets, D., Terzopoulos, V., **Jain, S.**, Sima, D.M., Struyfs, H., Engelborghs, S., and Van Hecke, W., ADmetrix: A new method for atrophy quantification in Alzheimer's disease, AAIC 2015, abstract and paper poster presentation, status-online.
12. Lysandropoulos, A., Ribbens, A., **Jain, S.**, Maertens, A., Van Hecke, W., Mavroudakis, N., Absil, J., Metens, T., and David, P., Comparison of 1.5 versus 3 Tesla MRI scanner in Quantification of Brain Atrophy in MS, Neurology 2015, abstract and paper poster presentation, status-online.
13. Sima, D.M., **Jain, S.**, Maertens, A., Van Vlierberghe, E., Van Hecke, W., and Smeets, D., MSmetrix validation of normative brain volume population graphs to serve as a reference in the clinical follow up of MS patients, ECTRIMS 2016, paper poster presentation, status-online.
14. Bouckaert, F., Emsell, L., Claes, L., Dewinter, F-L., **Jain, S.**, Dols A., Wampers M., Stek M., Stefan S., Sienaert P., and Vandenbulcke M., Hippocampal Volume, Serum BDNF and Clinical Changes Following Electroconvulsive Therapy in Late Life Depression, Society of Biological-Psychiatry on Stress, Emotion, Neurodevelopment and Psychopathology, Biological Psychiatry, 2015, status-online.

Patents

Jain, S., Smeets, D., Sima, D.M., Ribbens, A., and Maertens, A., Method and system for analyzing image data, EP Patent App. EP20140184117, 2016, status-online.

Thesis

Jain, S., Early diagnosis of Alzheimer disease, VIBOT, Heriot-Watt University (UK), June 2013.

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
ESAT-PSI
Kasteelpark Arenberg 10 box 2441
B-3001 Leuven
saurabh.jain@kuleuven.be
http://www.esat.kuleuven.be/psi